9-16-2014 12:00 AM

Psychometric evaluation of self-report pain and disability measures for elbow disorders

Joshua Vincent
The University of Western Ontario

Supervisor
Joy C. MacDermid
The University of Western Ontario

Graduate Program in Health and Rehabilitation Sciences
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

PSYCHOMETRIC EVALUATION OF SELF-REPORT PAIN AND DISABILITY MEASURES FOR ELBOW DISORDERS
(Thesis format: Integrated Article)

by

Joshua I. Vincent

Graduate Program in Health and Rehabilitation Sciences

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

ABSTRACT

Elbow disorders are one of the commonest musculoskeletal problems with a prevalence of 9% in men and 8.1% in women. Patient centered care is the goal of current healthcare delivery models; but optimizing treatment outcome and clinical research is hampered by a lack of outcome measures. Since pain and disability resulting from elbow disorders are experienced differently across individuals, they are best captured by patient reported outcome measures (PROM).

PROMs like the Patient-rated Elbow Evaluation (PREE); American Shoulder and Elbow surgeons – Elbow form (pASES-e) have been developed for use in elbow disorders, but important questions remain about their measurement properties. The key questions are: 1) what is the structural, construct validity and responsiveness of existing PROMs? 2) Does the PREE fit a continuous metric? 3) Do PROMs reflect the concerns that are important to patients and the World Health Organization's International Classification of Functioning Disability and Health (ICF) and its core sets? And finally, 4) what is known after synthesizing this new information with all prior knowledge on measurement elbow-related disability?

The overall objective of this thesis is to evaluate the psychometric properties of the PREE and the pASES-e. We used a mix of modern and traditional psychometric methods to assess the psychometric properties of the two elbow PROMs. We analyzed the construct validity, sensitivity to change, factor structure and internal consistency of these two measures using classical test methods. Then we synthesized the literature on psychometric properties of these two measures by conducting a systematic review. International Classification of Functioning Disability and Health (ICF) was used to analyze the content of these PROMs and compare them to the concerns self-nominated by patients with regards to functional activities. Finally Rasch analysis of the PREE was completed.

The results of the thesis indicate that the PREE and the pASES-e are valid reliable and sensitive to change. Both measures exhibited acceptable levels of content validity and have confirmed to the framework of the ICF. They also had enough depth and breadth to cover the important concerns related to function self-endorsed by elbow disorder patients. The PREE

satisfied Rasch model requirements with minimal data handling with a potential for obtaining an unbiased interval level estimate. Appropriate recommendations have been made for future research based on the outcomes of this work.

This work has enabled us to establish a core set of measures that are valid, reliable and sensitive to change to measure activity limitation and participation restrictions in people with elbow disorders that are critical to advancing clinical research and practice.

KEYWORDS: PREE, pASES-e, clinical measurement properties, validity, reliability, responsiveness, ICF, Rasch analysis

CO-AUTHORSHIP

Research question, specific objectives and individual study design were developed by Joshua I. Vincent and Joy C. MacDermid with inputs from Graham J.W. King and Ruby Grewal. Co-investigators were recruited when additional raters with specific expertise were required. Thesis advisory committee members were included as co-authors for specific chapters based on their input to individual chapters in this thesis work. The authors and specific roles for each chapter of the thesis are listed below:

CHAPTER 1: INTRODUCTION

Joshua I. Vincent – sole author

CHAPTER 2: VALIDITY AND SENSITIVITY TO CHANGE OF PATIENT-REPORTED PAIN AND DISABILITY MEASURES FOR ELBOW PATHOLOGIES

Joshua I. Vincent – primary author, study design, data analysis and wrote manuscript
Joy C. MacDermid – study design, data analysis and reviewed manuscript
Graham J.W. King – Study design, provided subjects and reviewed manuscript
Ruby Grewal - Study design, provided subjects and reviewed manuscript

CHAPTER 3**:** MEASUREMENT OF ELBOW PAIN AND FUNCTION: A SYSTEMATIC REVIEW OF THE PSYCHOMETRIC PROPERTIES OF TWO ELBOW SELF-REPORT MEASURES

Joshua I. Vincent – primary author, responsible for study design, literature review, quality appraisal, data extraction, narrative synthesis and manuscript writing
Joy C. MacDermid – study design, data analysis and reviewed manuscript
Graham J.W. King – Study design and reviewed manuscript
Ruby Grewal - Study design and reviewed manuscript
Emily Lalone – Quality appraisal, data extraction and reviewed manuscript

CHAPTER 4: LINKING OF THE PATIENT RATED ELBOW EVALUATION (PREE) AND THE AMERICAN SHOULDER AND ELBOW SURGEONS – ELBOW QUESTIONNAIRE (PASES-E) TO THE INTERNATIONAL CLASSIFICATION OF FUNCTIONING DISABILITY AND HEALTH (ICF) AND HAND CORE SETS

Joshua I. Vincent – primary author, responsible for study design, rater for ICF linking, data analysis and wrote manuscript
Joy C. MacDermid – study design, rater for ICF linking, data analysis and reviewed manuscript
Graham J.W. King – Study design and reviewed manuscript
Ruby Grewal - Study design and reviewed manuscript

CHAPTER 5: THE PATIENT RATED ELBOW EVALUATION AND THE AMERICAN SHOULDER AND ELBOW SURGEONS – ELBOW FORM CAPTURE ASPECTS OF FUNCTIONING THAT ARE IMPORTANT TO PATIENTS WITH ELBOW INJURIES

Joshua I. Vincent – primary author, responsible for design, data collection, data analysis, rater for ICF linking and wrote manuscript
Joy C. MacDermid – study design, rater for ICF linking, data analysis and reviewed manuscript
Graham J.W. King – Study design, provided subjects and reviewed manuscript
Ruby Grewal - Study design, provided subjects and reviewed manuscript

CHAPTER 6: RASCH ANALYSIS OF THE PATIENT RATED ELBOW EVALUATION

Joshua I. Vincent – primary author, responsible for study design, data organizing, Rasch analysis and wrote manuscript
Joy C. MacDermid – study design, Rasch analysis and reviewed manuscript
Graham J.W. King – Study design, provided subjects and reviewed manuscript
Ruby Grewal - Study design, provided subjects and reviewed manuscript

CHAPTER 7: GENERAL DISCUSSION

Joshua I. Vincent – sole author

ACKNOWLEDGEMENT

I would like to thank the Lord God Almighty for all the blessings that He has showered upon me all throughout my life, especially this special one.

I would like to thank my supervisor Dr. Joy C. MacDermid. Joy, words cannot describe how thankful I am to you for your continual unwavering support and guidance that has brought me to where I am today in my academic life. Still I remember the day, four years ago, when you talked on how to complete a manuscript in sixteen days. It was that spark that set me on fire for all the work that I was able to accomplish in the past four years. I have always admired the way you deal with volatile situations and have learnt lessons from them. I have always looked on to you as a mentor and as a guide. You told that you trust me with the work that I am doing and I hope that I stood up to your expectations.

I would like to thank my advisory committee members Dr. Ruby Grewal and Dr. Graham J.W. King for their advice and direction throughout the course of my PhD. Ruby, thank you for being a powerful role model and a mentor. I have always admired your keen interest on details in spite of your busy schedule. I will always be glad that we had that conversation at my desk, when I messed up at work, which really changed my outlook about relating to people and voicing my views without fear. Thank you for all the help that you have been in setting up my career as a researcher. Graham, I have finally included your middle initials correctly this time in your name, without you having to change it as you have done through these years. You have always been a down-to-earth person with subtle promptings towards the right direction. Thank you for all the help that you have been to me.

I would like to thank all my colleagues and lab mates especially Jayaprakash Raman, Kate Kelly and Derek Cheung. JP I am always grateful to you for the guidance that you gave me in choosing this university. Kate thanks for being an amazing co-worker and friend. I am thankful for all the help that you have been to me with this thesis. Derek, man you have been a great friend and an awesome lab mate. I will always remember the practical help that you were to me in settling down here; I always cherish your friendship. I would also like to thank all the surgeons, fellows, residents and nurses at the Roth McFarlane Hand and Upper Limb Clinic for their help in screening and recruiting patients for the study.

I take great pride in thanking my mom (Mrs. Kalyani Vincent), my sister (Ms.Janet Vincent) and my brother (Mr. Jaba Vincent). Mummy, Jem and Sam, because of you I am who I am today. You have been the propelling force that has made me stay in school since 2002. Mom along with Dad, you have instilled all the good values and the fear of God in me that has made me the man that I am today. And I know that I have brought you the joy and happiness that you deserve. We as a family have always been one single unit bound with cords of love during the highs and lows of life. And I know we will be like that until the end.

Finally, Vanitha my beloved wife, I have no words to thank you for all the help that you have been with this thesis and in my life at large. You have been my best friend and now you are the love of my life. I can promise you that all those times that I missed to spend with you is over; and I know you did not make a big deal of it. You have been very patient, enduring and hard working. And I dedicate this thesis to you along with my mom, sister and brother.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 6: RASCH ANALYSIS OF THE PATIENT RATED ELBOW EVALUATION

CHAPTER 1: INTRODUCTION

Elbow disorders and outcome assessment

Disorders of the elbow can range from simple tendinitis to malignant tumors.[11] Regardless of the conditions newer and better treatment options are available for management of these conditions.[1] This is made possible due to high quality research on the outcomes of these treatment procedures. Measurement of health outcomes is an integral part of clinical practice and health research.[2,] As for any other joint or condition, measuring outcomes of elbow disorders and its management have three main purposes: discrimination, prediction and evaluation.[3] Discrimination is the ability to differentiate between different levels of health and disability. Prediction is the ability of an outcome measure to predict future events. While, evaluation is the ability to measure the outcome of a medical, surgical or rehabilitative intervention.

Types of outcome measures

There are two broad categories of outcome measures: clinician based or performance based outcome measures (CBO) and patient rated outcome measures (PROM).[4] CBO measures are clinical tests or observations performed by a clinician to evaluate the status of a patient's health condition in a clinical setting. [4] A PROM is defined as any report of the status of a patient's health condition that comes directly from the patient, without interpretation by another person.[5] One main disadvantage of using the CBO is that they may not accurately capture the perspective of patient. Realizing the importance of PROMs several organizations including the Food and Drug Administration (FDA) have made the use of PROMs mandatory in drug label claims.[5]

In the last couple of decades, outcomes research has been influenced by the paradigm shift of importance from 'efficacy' to 'effectiveness'. Efficacy is defined as the biological effect of treatment delivered under carefully controlled conditions.[6] In contrast, effectiveness is defined as the usefulness of a particular treatment to the individuals receiving it under typical clinical conditions.[6] The emphasis has shifted from measuring narrow biological effects of treatments to assessing the broader biopsychosocial impact of health interventions.[3] In other words, it is the broadening of clinical outcomes research along a continuum from measurements at the level of body functions/ body structures to focusing on activity limitation and participation

restriction.[7,8] This paradigm shift has made this whole concept of measuring health and disability a complex matrix needing a conceptual framework to understand all the factors that come into play.

Models of health

There are various frameworks that help in guiding clinicians and health researchers in understanding what aspects of health ought to be measured. Conceptual models of particular relevance to rehabilitation include Nagi's model,[9] the International Classification of Impairments, Disabilities and Handicaps (ICIDH)[10] and the International Classification of Functioning Disability and Health (ICF).[11] The first two models are social models, with inherent limitations. They look at disability as a socially created problem and not the problem of the individual.[12] The ICF is a biopsychosocial model that integrates the medical and social models of disability.[11,13] Here disability is perceived as a consequence of biological, personal, and social dimensions.[14] The ICF model which is the current international framework for measuring health and health related states will be discussed in detail.

International classification of Functioning Disability and Health (ICF)

The ICF is a biopsychosocial model which provides a clearer synthesis of the previous models of disability thus providing a universal language to discuss functioning and disability across disciplines and borders. [11-15] It was accepted by the World Health Assembly in the year 2001. Within the ICF, the term 'health condition' is used to describe the health problem. The ICF has two major parts: a conceptual model and a coding system. The conceptual model helps us to understand the dynamic interaction between a person's health condition and other contextual factors. (See Figure 1) The conceptual model has 2 main parts – functioning and disability; and contextual factors. The functioning and disability part has three domains: body structures and body functions - depicting functioning at the level of body parts; activities – depicting functioning at the level of a whole person; and finally participation – depicting functioning of a whole person in their complete environment. The contextual factors part includes environmental factors and personal factors.[11,12,14] In the model the domains are connected by bi-dimensional arrows indicating interaction between the domains.[13]

The ICF coding system increases its utility in clinical practice and research.[15-17] The coding system is an hierarchical one consisting of 1440 alpha numeric codes divided into four main domains of Body Functions ('b' codes - 493), Body Structures ('s' codes - 310), Activities and participation ('d' codes - 384) and Environmental factors ('e' codes - 253). The codes are organized into components, chapters (1st level) and categories on 2nd, 3rd and 4th level. The ICF code d6301 – preparing complex meals is used as example to describe the various levels. Here'd' indicates the activity and participation domain; '6' indicates the chapter level – here it is domestic life; '3' indicates 2nd level category; '0' indicates 3rd level category and '1' indicates 4th level category. The higher the level used more specific is the detail that is described by the code.[15,16,17]

ICF core set for Hand conditions

To make the application of ICF more feasible in clinical practice and research core sets for specific conditions have been developed. These core sets are a subset of ICF categories that are relevant to a particular disease or condition. There are two types of core sets, a comprehensive core set and a brief core set. The comprehensive core set consists of those ICF categories that make a comprehensive and exhaustive description of the condition.[18] While the brief core set consists of the most essential categories that can serve as a minimal standard for describing the condition.[18] There are currently 34 core sets that have been developed for various conditions including osteoporosis; amputees; chronic pain; rheumatoid arthritis; osteoarthritis; ankylosing spondylitis; hand conditions etc.[18]

The core set for Hand conditions was developed by the ICF research branch to provide a common language for clinicians who are specialized in hand conditions to assess patients with a variety of hand disorders. It was approved by an international panel of experts at a consensus conference.[19] The comprehensive core set for hand conditions consists of 117 codes and the brief core set contains 23 codes.[19,20] The published document on the conference refers to hand conditions as any problem that is directly located at the hand. It may range from carpel tunnel syndrome to amputation; and also conditions located in different parts of the body with a manifestation at the hand. As such there is no core set that is available to assess elbow conditions. Since the upper extremity (shoulder, elbow and wrist) functions as a coordinated unit it may be hypothesized that the core set used for hand conditions may address functional

concerns of patients with elbow disorders. The hypothesis is then it would increase clinical utility of the core set as clinicians could use one common core set elbow, wrist and hand conditions rather than having individual core sets for these joints.

Classification of Elbow outcome measures based on the ICF framework

The ICF provides a universal framework for clinicians to decide on what they should measure. [21] It has been suggested that body functions and body structures can be measured by objective measures such as imaging and physiological tests; activity and participation can be measured by joint or region specific patient based questionnaires and environmental and personal factors can be comprehensively measured by generic patient reported measures of quality of life and biopsychosocial measures of emotions, psychological status or social concerns.[22] Based on a review of the literature [23] we have created a table to show the outcome tools available to measure the different domains of the ICF in general elbow disorders. (See table 1) There have also been previous studies that have provided ICF based clinical guidelines as to what should be measured and the instruments that are available to measure the needed constructs in other conditions like stroke, rheumatoid arthritis etc.[7,21] Liem and colleagues [22] created a framework for outcome measurement in general elbow disorders. They have suggested the use of a generic PROM to measure quality of life; a regional PROM to measure upper extremity function; an elbow-specific PROM to measure elbow function; and observer based clinician data for objective assessment. They recommended the use of Patient Rated Elbow Evaluation (PREE) for the elbow-joint specific PROM to measure function. [22] In a similar study Angst and colleagues created a framework for assessment in elbow arthroplasty, and suggested that the PREE and the self-report section of the American Elbow and Shoulder Surgeons – Elbow form (pASES-e) can be used as elbow specific PROMs.[24]

Patient Reported Outcome Measures for elbow conditions

Since PROMs complement rather than replace CBO, and provide a patient centered assessment of function, it is important to ensure the measurement rigor of PROM designed for elbow conditions. The two commonly used elbow pain and disability measures are the Patient Rated Elbow Evaluation form (PREE) and the self-report section of the American Elbow and Shoulder Surgeons – Elbow form (pASES-e).

The Patient Rated Elbow Evaluation form (PREE):

The PREE is a 20 item PROM, consisting of two main sections, pain and function. [25] Responses are rated on a 0 to 10 numeric rating scale. The pain section consists of five items that quantify the severity of pain in different situations and the frequency of pain. The function section has 15 questions divided into two sub sections usual activities and specific activities. In this section patients are asked to rate their difficulty in performing activities of daily living ranging from simple activities like getting dressed up to complex roles like work. All the scores are computed to obtain a global score out of 100. Higher PREE total scores reflect greater pain and disability.[25] It takes around 3 minutes to complete the PREE. The PREE has been translated into Japanese[26] and German.[27]

The psychometric properties of the PREE have been evaluated using traditional psychometric methods. It has demonstrated acceptable levels of construct validity. MacDermid (2001)[25] in the developmental study of the PREE has reported that PREE had exhibited Pearson's correlations ranging from 0.49 to 0.84 with the Disabilities of the Arm Shoulder and Hand questionnaire (DASH), pASES-e and the physical component summary score of SF-36. The PREE has been found to have a very high level of test retest reliability with an ICC of 0.95.[25] For the Japanese[26] and the German version[27] the ICCs ranged from 0.73 to 0.94. Hanyu et al[26] and John et al[27] reported a cronbach alpha 0.97 and 0.96 respectively for the Japanese and the German version. In a study Angst et al (2005)[28] analyzed the factor structure of the PREE along with the pASES-e and the DASH and have reported three principal components explaining 89.2% of variance, of which the component 'physical specific' to which the PREE items loaded the maximum explains 60.1% of variance. They used the total scores of the three PROMs and did not use a traditional approach to exploratory factor analysis. Responsiveness has been evaluated in a sample of total elbow arthroplasty patients and has been found to be acceptable (ES = 1.5; SRM = 1.37).[29] In summary, the PREE has demonstrated acceptable levels of psychometric properties based on the traditional methods. Gaps have been identified in the psychometric evaluation efforts that have been taken before. The PREE to date has not been subjected to modern psychometric methods like Rasch analysis[30] or item response theory (IRT)[31]

The self-report section of the American Elbow and Shoulder Surgeons – Elbow form (pASES-e):

The pASES-e is an 18 item PROM consisting of three sections pain, function and satisfaction.[32] The pain section has five questions and is scored on a 0 to 10 scale similar to that of the PREE. The function section measures function on both the affected and the unaffected elbow. A 0 – 3 ordinal scale is used to rate the responses. It consists of 12 items measuring various functional activities ranging from feeding to work and recreation. The satisfaction section has one item asking patients to rate their satisfaction with treatment on a 0 – 10 scale. Global score is not calculated for the pASES-e usually. For a comprehensive view of patient's condition the ASES also has a clinician part (cASES-e). It takes around 3 minutes to complete the pASES-e.[29] The pASES-e has been translated into German.[33]

The pASES-e has been evaluated using traditional psychometric methods. Acceptable levels of construct validity have been demonstrated by the pASES-e with other measures measuring a similar trait. In the PREE developmental study the pASES-e exhibited high to moderate correlations with the DASH, PREE and the SF-36 physical component summary score (Pearson's correlation 0.51 to 0.85).[25] In a study by Angst and colleagues[28] the pASES-e correlated highly with the PREE (Spearman's rho 0.92) and moderately with the DASH (Spearman's rho 0.73).[28] The pASES-e has demonstrated acceptable levels of test re-test reliability. MacDermid (2001)[25] have found the pASES-e pain subscale (ICC=0.89); the pASES-e function affected sub scale (ICC=0.79) and the pASES-e function unaffected sub scale (ICC=0.64) to demonstrate acceptable levels of test-retest reliability. Turchin and colleagues[34] have reported that the pASES-e exhibit a reliability of 0.79 in a sample of 69 patients with elbow pain due to various causes. The English version's internal consistency has not been evaluated. The German version exhibited a Cronbach's alpha of 0.90.[33] The factor structure of the pASES-e has not been analyzed traditionally but the total score was put into a model along with the PREE and the DASH. The pASES-e loaded onto the physical specific component along with the PREE that accounted for 60.1% of the variance.[28] The responsiveness of the pASES-e has been evaluated in total elbow arthroplasty patients. The pASES-e has demonstrated acceptable levels of ES and SRM (ES = 1.32; SRM = 1.17).[29] In summary, the pASES-e has demonstrated

acceptable levels of psychometric properties. However there are some gaps in the literature. The pASES-e has never been subjected to modern psychometric analysis. The longitudinal validity, factor structure and responsiveness have not been studied in detail.

Criteria for evaluation of patient reported outcome measures

Many guidelines have been developed that need to be used while selecting an outcome measure.[2,4,21,35,36] The International Society for Quality of Life Research (ISOQOL) has recommended the minimum standard for PROMs.[35] This includes documentation of the conceptual and measurement model; evidence for reliability, validity (content validity, construct validity, responsiveness); interpretability of scores; quality translation, and acceptable patient and investigator burden.[35] Definitions of important clinical measurement properties are provided in Table 2. It would be useful for a clinician to have all this information so as to make informed evidence based decision on which outcome measure that he or she can use in routine clinical practice.

Reliability is the degree to which a PROM is free from measurement error.[37] There are two kinds of reliability: 1) test retest reliability (stability) and 2) internal consistency (homogeneity). Test retest reliability is the ability of a measure to provide consistent scores over time in a stable population.[38,39] Test-retest reliability is usually measured by calculating Intra class correlation coefficient (ICC).[40] The minimum standard recommended is an ICC of 0.90 for use in a clinic setting.[40] Internal consistency also called homogeneity, is the measure of inter relatedness of the items that are included in a PROMs.[41] It is measured by calculating Cronbach's alpha.[42] Cronbach's alpha should range from 0.70 to 0.95.[40] Modern psychometric methods like item response theory (IRT)[31] or Rasch analysis[30] can also be used as a measure of reliability.

Validity is the degree to which a PROM measures the construct it purports to measure.[39] There are three broad types of validity. They are content validity; construct validity; criterion validity. Content validity is the extent to which the PROM represents the most relevant and important aspects of a concept in a given context.[39] This is a very important criterion for any PROM. There are many different ways of establishing content validity. One is through expert and patient consensus. This can be done through surveys, focus groups, interviews etc. The latest

in the methods to evaluate content of PROMs is the linking of the items of the PROM to the ICF using the ICF codes following standardized linking rules.[15,16]

Construct validity is the degree to which the scores on the PROM relate to other measures (both patient reported and impairment based) measuring similar construct in a manner that is consistent with theoretically derived apriori hypotheses concerning the concepts that are being measured. [39] It is recommended that for an instrument to have sufficient construct validity empirical evidence to support predefined hypotheses on the expected associations among measures similar or dissimilar to the measured PROM should be presented.[35] It is usually measured by calculating Pearson's correlation coefficient (*r)* or Spearman's rank correlation. Moderate to large correlations (>0.40) between related constructs are expected to support construct validity. [35]

Responsiveness is the ability of a PROM to capture changes over time in the construct being measured. [37,43,44] It is an important clinical measurement property, as a measure that is not sensitive to change cannot capture the change in patients' health status resulting from a therapeutic, medical or surgical procedure. There are two major methods of measuring responsiveness anchor based and distribution based. In the anchor based method a tool to measure change like the Global Rating of Change scale (GRoC) is administered to patients to rate the change in their condition. Then a minimal clinically important difference (MCID) is calculated which is a very important indicator of change in health condition in response to treatment that is perceived as important by clinicians and patients. [45,46] In the distribution based method the mean change is used to calculate effect sizes (ES) and standardized response means (SRMs). [37,43] These two indices help in capturing true change. This may or may not be perceived as important by patients.

There are other factors such as interpretability of scores, administrative burden and patient burden that need to be considered when selecting a PROM. Interpretability is the ease with which a score obtained from a PROM can be assigned a meaning. [2,47]Administrative burden is an umbrella term that includes many factors like time needed to administer a questionnaire, training, cost, time taken to calculate the score etc.[2] Patient or respondent burden include the time needed to fill in a PROM, ease of understanding the questions etc. [2]

Gaps in current knowledge

Based on our literature review and by comparing the literature on the clinical measurement properties of the PREE and pASES-e against minimal standards as set by ISOQOL we found the following gaps in the literature.

1) The content of the PREE and the pASES-e have never been formally assessed for validity. It is always the best practice to know the content of the PROM that a clinician is going to use. Another aspect of evidence based practice is to see if the items cover concepts that are important to patients.

2) The factor structure of the PREE and the pASES-e have not been sufficiently explored

3) Responsiveness of these two PROMs has been studied only in total elbow patients and has not been studied in other elbow disorder population.

4) The PREE nor the pASES-e have ever been subjected to modern psychometric methods like Rasch analysis and Item response theory (IRT).

5) There is no summary document synthesizing all the clinical measurement properties of the PREE and the pASES-e.

Research question

Do elbow patient reported outcome measures of pain and disability exhibit sufficient psychometric properties?

Objectives

The overall objective of this thesis is to evaluate the psychometric properties of the Patient Reported Elbow Evaluation and the self-report section of the American Shoulder and Elbow Surgeons – Elbow form. Based on the research gaps identified in the previous sections the specific objectives are as follows:

1) To evaluate the internal consistency, concurrent construct validity, longitudinal validity, sensitivity to change and factor structure of the PREE, the pASES-e and the DASH in a diverse group of patients surgically managed for various elbow pathologies.

2) To perform a comprehensive systematic review of the literature and summarize the quality and the content of the evidence that is available on the psychometric properties of the PREE and the pASES-e.
3) To analyze the conceptual basis of the PREE and the ASES-e by linking the meaningful concepts in these PROs to the ICF using standardized linking rules and to determine the extent to which the ICF core set for hand conditions cover the content of elbow questionnaires using summary ICF linkage indicators.
4) To use the ICF classification to summarize and identify aspects of functioning that are reported as important by a cohort of patients with elbow disorders and compare it to the content of the PREE and the pASES-e.
5) To conduct a rasch analysis of the PREE to assess the overall fit to the rasch model, the response scale used, individual item fit, differential item functioning (DIF), local dependency, unidimensionality and person separation.

Thesis Overview

Chap 2 describes the psychometric evaluation performed on the PREE and the pASES-e to analyze the internal consistency, construct validity, longitudinal validity, sensitivity to change and factor structure. The DASH was used as the comparator in this study. This is in line with the first objective. To fulfill our second objective, a systematic review of the psychometric properties of the PREE and the pASES-e are presented in chap 3. The next three chapters are set up based on the gaps identified and the recommendations made in the systematic review. Chap 4 describes the content analysis of the PREE and the pASES-e performed using the ICF framework and the codes contained in it. This is to satisfy objective 3. Objective 4 of the thesis is accomplished by the study presented in chap 5 where activities involving the elbow that were important to patients were identified and then the content of the PREE and the pASES-e were compared to the patient self-endorsed concerns to see if they capture them. The ICF was used as a standard framework to enable this comparison. The last chapter of this thesis uses a modern clinical measurement methodology called the rasch analysis to identify and correct the factors that prevent clinicians and researchers from obtaining a linear interval level score from the PREE. This is in line with objective 5. Chap 7 provides a discussion and overview of the thesis

work. Also it discusses the strengths, limitations, clinical and research implication, and future directions of this thesis work.

REFERENCES

1. Morrey BF. The elbow and its disorders, 4th ed. *Scitech Book News*. 2009;33(2)

2. Aaronson N, Alonso J, Burnam A, et al. Assessing health status and quality-of-life instruments: Attributes and review criteria. *Qual Life Res*. 2002;11(3):193-205.

3. Cohen ME, Marino RJ. The tools of disability outcomes research functional status measures. Arch Phys Med Rehabil. 2000;81, Supplement 2(0):S21.

4. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess*. 1998;2(14):i-iv, 1-74.

5. Food and Drug Administration. Patient reported outcome measures: Use in medical product development to support labeling claims. 2009; Accessed 11/8/2011.

6. Flay BR. Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Prev Med*. 1986;15(5):451-474.

7. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J. Issues for selection of outcome measures in stroke rehabilitation: ICF body functions. *Disability & Rehabilitation*. 2005;27(4):191-207.

8. Rosenbaum Peter D,Stewart. The world health organization international classification of functioning, disability, and health: A model to guide clinical thinking, practice and research in the field of cerebral palsy. *Seminars in pediatric neurology*. 2004;11(1):5-10.

9. Nagi H. Some conceptual issues in disability and rehabilitation. In: Sussman MB, ed. *Sociology and rehabilitation.* Washington DC: American Sociological Association; 1965:100-113.

10. World Health Organization (WHO). *International classification of impairments, disabilities, and handicaps: A manual of classification relating to the consequences of disease.* Geneva, Switzerland: World Health Organization; 1980.

11. World Health Organization (WHO). *The international classification of functioning disability and health (ICF).* Geneva, Switzerland: World Health Organization; 2001.

12. Jette AM. Toward a common language for function, disability, and health. *Phys Ther*. 2006;86(5):726-734.

13. Cieza A, Stucki G. The international classification of functioning disability and health: Its development process and content validity. *Eur J Phys Rehabil Med*. 2008;44(3):303-313.

14. Stucki G, Ewert T, Cieza A. Value and application of the ICF in rehabilitation medicine. *Disabil Rehabil*. 2002;24(17):932-938.

15. Cieza A, Brockow T, Ewert T, et al. Linking health-status measurements to the international classification of functioning, disability and health. *J Rehabil Med*. 2002;34(5):205-210.

16. Cieza A, Geyh S, Chatterji S, Kostanjsek N, Ustun B, Stucki G. ICF linking rules: An update based on lessons learned. *J Rehabil Med*. 2005;37(4):212-218.

17. Cieza A, Stucki G. Content comparison of health-related quality of life (HRQOL) instruments based on the international classification of functioning, disability and health (ICF). *Quality of Life Research*. 2005;14(5):1225-1237.

18. Bickenbach J, Rauch A, Stucki G. *ICF core sets. manual for clinical practice.* Hogrefe, Göttingen: Hogrefe & Huber Pub; 2012.

19. Rudolf K, Kus S, Chung KC, Johnston M, LeBlanc M, Cieza A. Development of the international classification of functioning, disability and health core sets for hand conditions â€“ results of the world health organization international consensus process. *Disabil Rehabil*. 2012;34(8):681-693.

20. Selb M, Escorpizo R, Kostanjsek N, Stucki G, Ustun B, Cieza A. A guide on how to develop an international classification of functioning, disability and health core set. *Eur J Phys Rehabil Med*. 2014.

21. Kucukdeveci AA, Tennant A, Grimby G, Franchignoni F. Strategies for assessment and outcome measurement in physical and rehabilitation medicine: An educational review. *J Rehabil Med*. 2011;43(8):661-672.

22. Liem IS, Kolling C, Marks M, Nelissen RG, Goldhahn J. Development of a score set to measure function and quality of life in patients suffering from elbow pathology. *Arch Orthop Trauma Surg*. 2012;132(6):831-837.

23. Longo UG, Franceschi F, Loppini M, Maffulli N, Denaro V. Rating systems for evaluation of the elbow. *British Medical Bulletin*. 2008;87:131-161.

24. Simmen BR, Angst F, Schwyzer HK, et al. A concept for comprehensively measuring health, function and quality of life following orthopaedic interventions of the upper extremity. *Arch Orthop Trauma Surg*. 2009;129(1):113-118.

25. MacDermid JC. Outcome evaluation in patients with elbow pathology: Issues in instrument development and evaluation. *Journal of Hand Therapy*. 2001;14(2):105-114.

26. Hanyu T, Watanabe M, Masatomi T, et al. Reliability, validity, and responsiveness of the japanese version of the patient-rated elbow evaluation. *Journal of Orthopaedic Science*. 2013;18(5):712-719.

27. John M, Angst F, Pap G, Junge A, Mannion AF. Cross-cultural adaptation, reliability and validity of the patient rated elbow evaluation (PREE) for german-speaking patients. *Clin Exp Rheumatol*. 2007;25(2):195-205.

28. Angst F, John M, Pap G, et al. Comprehensive assessment of clinical outcome and quality of life after total elbow arthroplasty. *Arthritis Care and Research*. 2005;53(1):73-82.

29. Angst F, Goldhahn J, Drerup S, et al. Responsiveness of five outcome measurement instruments in total elbow arthroplasty. *Arthritis Care and Research*. 2012;64(11):1749-1755.

30. Andrich D. *Rasch models for measurement.* Vol 07-068. Newbury Park: Sage Publications; 1988:95.

31. DeMars C. *Item response theory.* Oxford ;; New York: Oxford University Press; 2010:131.

32. King GJ, Richards RR, Zuckerman JD, et al. A standardized method for assessment of elbow function. research committee, american shoulder and elbow surgeons. *J Shoulder Elbow Surg*. 1999;8(4):351-354.

33. John M, Angst F, Awiszus F, King GJ, MacDermid JC, Simmen BR. The american shoulder and elbow surgeons elbow questionnaire: Cross-cultural adaptation into german and evaluation of its psychometric properties. *J Hand Ther*. 2010;23(3):301-13.

34. Turchin DC, Beaton DE, Richards RR. Validity of observer-based aggregate scoring systems as descriptors of elbow pain, function, and disability. *J Bone Joint Surg Am*. 1998;80(2):154-162.

35. Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*. 2013;22(8):1889-1905.

36. Lohr KN, Aaronson NK, Alonso J, et al. Evaluating quality-of-life and health status instruments: Development of scientific review criteria. *Clin Ther*. 1996;18(5):979-992.

37. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. statistics and strategies for evaluation. *Control Clin Trials*. 1991;12(4 Suppl):142S-158S.

38. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: What are the necessary measurement properties? *J Clin Epidemiol*. 1992;45(12):1341-1345.

39. Portney L, Watkins M. *Foundation of clinical research: Applications to practice.* Upper Saddle River, NJ: Prentice-Hall; 2000.

40. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30-46.

41. Cortina JM. What is coefficient alpha? an examination of theory and applications. *J Appl Psychol*. 1993;78(1):98-104.

42. Zinbarg RE, Revelle W, Yovel I, Li W. Cronbach's α, Revelle's β, and Mcdonald's ωH: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*. 2005;70(1):123-133.

43. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: A critical review and recommendations. *J Clin Epidemiol*. 2000;53(5):459.

44. Beaton DE. Understanding the relevance of measured change through studies of responsiveness. *Spine (Phila Pa 1976)*. 2000;25(24):3192-3199.

45. Beaton DE, Bombardier C, Katz JN, et al. Looking for important change/differences in studies of responsiveness. OMERACT MCID working group. outcome measures in rheumatology. minimal clinically important difference. *J Rheumatol*. 2001;28(2):400-405.

46. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407-415.

47. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63(7):737-745.

Table 1: Elbow outcome measures classification based on ICF

ICF DOMAIN	ASSESSMENT DOMAIN	OUTCOME	OUTCOME MEASURE
Body functions	*Mobility of joint functions*	Joint range of motion, joint stiffness and Crepitus	• Flynn criteria • Neviaser criteria • Jupiter criteria; • MEPI • cASES-e • HSS scoring system • Ewald scoring system • Liverpool elbow score
	Muscle power functions	Grip Strength and Pinch strength	• Manual (JAMAR) or computerized (J tech) • HSS scoring system • Broberg and Morey rating system • Liverpool elbow score
	Sensation of touch	Two point discrimination	• Manual or computerized - two point discriminator
	Sensation of pain	Pain	• Visual analog scale • Numeric rating scale • PREE – pain section • pASES-e - pain section
	Tenderness	Pressure pain sensitivity	• Algometer • objective assessment by physician • cASES-e
	Stability of joints	Verbal and objective physician assessment	• MEPI • cASES-e • Broberg and Morey rating system

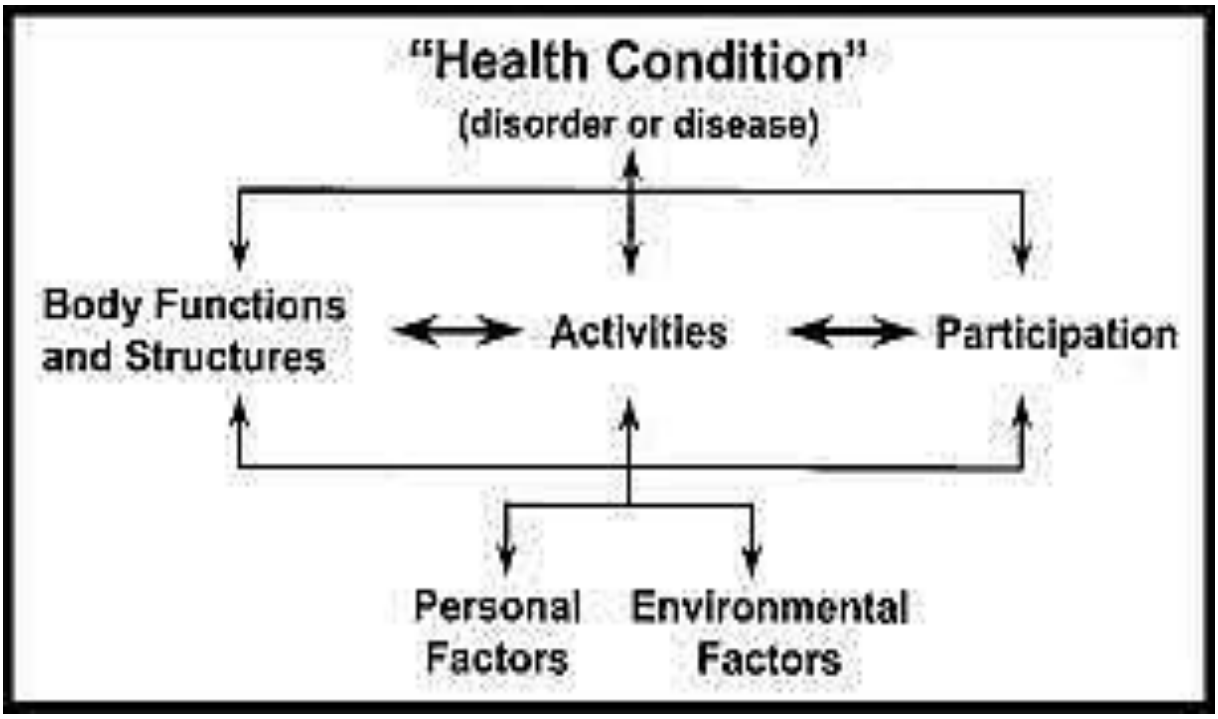
Body structures	*Structures related to movement*	Flexion contractures; and deformities (Valgus and Varus)	• Visual inspection • Goniometry • Radiograph • Ewald scoring system • cASES-e • HSS
	Structures related to strength	Muscular atrophy	• Visual observation • cASES-e
	Structure related nerve supply	Ulnar nerve compression	• Tinels sign • Liverpool elbow score • cASES-e
Activities and participation	*Performing everyday activities and societal roles*	Activity limitation and participation restriction (Usually self-reported)	• PREE, • pASES-e; • DASH; • Quick DASH; • Oxford elbow score (self-report section); • Liverpool elbow score (self-report section); • MEPI (self-report section); • SMFA

MEPI – Mayo Elbow Performance Index; **PREE** – Patient Rated Elbow Evaluation; **pASES-e** – Patient reported section of the American Shoulder and Elbow Surgeons – Elbow form; **cASES-e** – Clinician section of the American Shoulder and Elbow Surgeons – Elbow form; **HSS** – Hospital Scoring System; **DASH** – Disabilities of the Arm Shoulder and Hand questionnaire; **SMFA** – Short Musculoskeletal Functional Assessment

Table 2: Definitions of clinical measurement properties

CLINICAL MEASUREMENT PROPERTY	DEFINITION
Test-retest reliability	The degree to which an instrument is stable based on repeated administrations of the test to the same individual over a specified time interval.
Standard error of measurement	A reliability measure of response stability, estimating the standard error in a set of repeated scores.
Internal consistency	A form of reliability, assessing the degree to which a set of items in an instrument all measure the same trait.
Construct validity	A type of measurement validity; the degree to which a theoretical construct is measured by an instrument.
Convergent validity	An approach of construct validation, assessing the degree to which two different instruments or methods are able to measure the same construct.
Discriminant or divergent validity	An approach in construct validation assessing the degree to which an instrument yields different results when measuring two different constructs; that is, the ability to differentiate between the constructs.
Known group validity	A technique for construct validation, in which validity is determined by the degree to which an instrument can demonstrate different scores for groups known to vary on the variable being measured.
Responsiveness	The ability of a test to demonstrate change.
Factor analysis	An exploratory multivariate statistical technique used to examine the structure within a large set of variables and to determine the underlying dimensions that exist within that set of variables.

Figure 1: International Classification of Functioning Disability and Health conceptual model



Source: World Health Organization (WHO)

CHAPTER 2. VALIDITY AND SENSITIVITY TO CHANGE OF PATIENT-REPORTED PAIN AND DISABILITY MEASURES FOR ELBOW PATHOLOGIES[1]

ABSTRACT

Study design: Prospective Cohort Study

Background: Measuring functional outcomes after surgical procedures of the elbow requires valid patient-report pain and disability questionnaires. The Patient Rated Elbow Evaluation form (PREE), the patient-reported form of the American Shoulder and Elbow Surgeons Questionnaire elbow form (pASES-e) and the Disabilities of the Arm, Shoulder and Hand questionnaire (DASH) are commonly used questionnaires. There is insufficient evidence available concerning their validity and sensitivity to change.

Purpose: To evaluate the internal consistency, concurrent construct validity, longitudinal validity, sensitivity to change and factor structure of the PREE, the pASES-e and the DASH in a diverse group of patients surgically managed for various elbow pathologies.

Methods and analysis: Data were prospectively collected from 128 post elbow surgery patients (Mean age = 46.53; S.D. = 12.77). Patients completed the PREE, the pASES-e, the DASH and the SF-36 at baseline and 6 months post-surgery. Concurrent construct validity, longitudinal validity, sensitivity to change and factor structure were analyzed.

Results: Concurrent construct validity was demonstrated by confirmation of expected relationships; the strongest correlations were observed between the PREE pain score, the PREE total score, the pASES-e pain score and the DASH score (r = 0.73 to 0.87). The pASES-e function score correlated the least with other constructs. Longitudinal validity demonstrated a similar trend; the pASES-e pain change score and PREE change score were most strongly

---

[1] Reproduced with permission from Vincent JI, MacDermid JC, King GJ, Grewal R. Validity and sensitivity to change of patient-reported pain and disability measures for elbow pathologies. *J Orthop Sports Phys Ther*. 2013;43:263-274.http://dx.doi.org/10.2519/jospt.2013.4029. Copyright ©*Journal of Orthopaedic & Sports Physical Therapy®*

correlated; while the pASES-e function change score and DASH change score were moderate to weakly correlated. All 3 patient-report questionnaires demonstrated a large effect size (ES) and standardized response means (SRM) (>1.0). Structural validity was supported for the PREE (R^2=77.2%; 4 factors) and the pASES-e (R^2=74.4%; 4 factors); but not for the DASH (R^2=71.3%; 5 factors).

Conclusion: The PREE, the pASES-e and the DASH have acceptable validity and sensitivity to change. pASES-e function sub-scale is least sensitive to change and less correlated to other measures. *J Orthop Sports Phys Ther 2011*

Key Words*:* *elbow questionnaires; outcome measures; quality of life; PREE; pASES-e; DASH; SF-36.*

\

INTRODUCTION

Scoring systems to rate the extent of pain and disability have become an integral part of modern orthopedic practice. Two basic types of measures are in common usage. One type is clinician based outcome measures (CBO) and the other is the patient reported outcome questionnaires (PRO). [9, 33] CBO can be affected by observer bias [24] and may not reflect patient priorities; hence the emergence of patient-report measures. The debate over patient-report versus clinician-based measures has largely been resolved in that these measures provide different perspectives and are both needed. Patient-report measures provide patient-centered perspective;[34, 36] and tend to be predictive of participation outcomes like return-to-work.[28] Clinicians are increasingly using such measures to contribute to clinical decision-making. The validity of the measures, and their subscales; as well as their ability to detect clinical change are important considerations when choosing between the available measures. The Patient Rated Elbow Evaluation form (PREE),[25, 27] the patient-reported form of the American Shoulder and Elbow Surgeons questionnaire elbow form (pASES-e)[19] and the Disabilities of the Arm, Shoulder and Hand questionnaire (DASH)[14] are the commonly used PROMs in the management of orthopedic elbow disorders.

The body of literature on the elbow specific measures is small and has substantial gaps. The DASH has a large body of literature supporting its use in the shoulder and hand disorders; but few studies have specifically focused on elbow disorders. In the PREE developmental study[27] which included 70 subjects with various elbow pathologies treated surgically and non-surgically, it was found that the PREE (ICC=0.95); the pASES-e pain subscale (ICC=0.89); the pASES-e function affected sub scale (ICC=0.79), the pASES-e function unaffected sub scale (ICC=0.64) and the DASH (ICC=0.93) demonstrated acceptable levels of test-retest reliability. Turchin and colleagues have also reported DASH to exhibit excellent reliability (ICC=0. 92) in a sample of 69 patients with elbow pain due to various causes;[39] but found the modified ASES-e to exhibit slightly lower reliability of 0.79. There are no reports on the internal consistency of the PREE and the ASES-e; or for the DASH in an elbow sample.

With regards to construct validity, Angst et al[1] found that all these three measures demonstrated acceptable correlations in a sample of total elbow arthroplasty cases: the PREE and the ASES-e showed high correlations (Spearman's rho 0.92), while the DASH correlated

moderately with the PREE (Spearman's rho 0.68) and the ASES-e (Spearman's rho 0.73). MacDermid[27] also evaluated the construct validity of these three questionnaires and found them to be good, satisfying the hypotheses framed a priori (ASES pain vs PREE pain 0.93; ASES function vs PREE function -0.61; DASH vs elbow questionnaires 0.85; and moderate correlations to SF-36 physical component scores (-0.56); versus low correlation to SF-36 mental component scores (-0.23). There are no published reports on the longitudinal validity of the PREE and the pASES-e; or of DASH in an elbow sample.

Structural validity of individual scales is usually assessed by factor analysis. Typically, factor analysis informs our understanding of how items on a measure fit together into separate constructs (factors). Angst et al[1] used factor analysis in a different way; to identify how different scales represented constructs. He examined the PREE, the DASH, the pASESe and SF-36 and found three main domains explaining 89.2% of the variance of the instruments' main scores. The important limitation of their study is that they did not do a traditional factor analysis at the item level but rather examined entire measures as factors to show how constructs fit together. There are no other published reports of factor analysis of the PREE, pASES-e or the DASH in an elbow sample. Patient for elbow disorders have not been subjected to exploratory factor analysis; and this evidence is needed to determine if the proposed subscale structures are warranted.

There has been a debate in the literature about the dimensionality of DASH as evident from studies that sampled patients with disorders of the shoulder and wrist. [21] In the DASH development process, the dimensionality of the DASH was studied using a principal component analysis which revealed one component explaining 57% of the variance which was interpreted as to support the unidimensionality of DASH.[35] However a later study[21] on 991patients who underwent rehabilitation for various upper extremity disorders which included exploratory factor analysis followed by confirmatory factor analysis and Rasch analysis found the DASH to fit in a 3 factor model. These authors suggested that the results were inconclusive and indicated a need for further research. A review of the entire body of research addressing factor analysis of the DASH would require a systematic review. However, even a cursory review indicates a lack of consensus on DASH dimensionality. In this study, where an exploratory approach is needed for

the previously unexamined elbow scales, an exploratory factor analysis of the DASH is warranted for comparative purposes.

Given the current state of knowledge, it is clear that elbow-specific patient-report measures have been insufficiently examined and the support of their use in practice remains precarious until gaps in our knowledge are addressed. The purpose of this study was to describe the following in a diverse group of patients surgically managed for various elbow pathologies:

i) The internal consistency, concurrent construct validity, longitudinal validity and sensitivity to change of the PREE, the pASES-e and the DASH

ii) The factor structure of the PREE, the pASES-e and the DASH.

MATERIALS AND METHODS

Study Design

Prospective Cohort Study

Subjects

Data were prospectively collected from 128 patients aged between 21 and 79 years (Mean 46.5; SD 12.7) who had undergone a variety of surgical procedures of the elbow at the Hand and Upper Limb Centre of St Joseph's Hospital in London, Ontario. The cohort included biceps tendon repair (n= 62), radial head fixation (n=30) and radial head arthroplasty (n=36). Patients were included if they were aged between 20 and 80 years; and were able to read and write English. They were excluded if they were cognitively impaired; if they had neurological or psychiatric disorders; if they had any history of malignancy.

Patients signed consent forms in compliance with the approval provided by the Western University Health Sciences Research Ethics Board. Approximately two third of participants were male 84/128 (65.62%) and 118/128 (92.2%) were right hand dominant. The dominant and the non-dominant elbows were equally affected in this sample. Twenty-four (18.7%) of the participants were receiving worker's compensation (See Table. 1). Participants were asked to complete the PREE, the pASES-e, the DASH and the SF-36 at baseline and again at six months

post-surgery. Six-months was chosen as an ideal time to evaluate these questionnaires as variability in outcome scores would be present; but patients would not have achieved their final surgical outcome.

Outcome Measures

Patient-Rated Elbow Evaluation: The Patient-Rated Elbow Evaluation (PREE) [25, 27] is a 20 item patient reported outcome questionnaire that measures elbow-related pain and disability of the affected upper extremity. There is a pain subscale; and 2 function subscales- one that addresses specific activities, the other usual role performance. All the items of the questionnaire are scored on a 0-10 numeric rating scale. The section dealing with pain has five questions of which four of them rate pain from 'no pain' (0) to 'worst ever' (10). The fifth question rates how often the patient has pain with responses ranging from 'never' (0) to 'always' (10). The responses on the sub-scales for the function section range from 'no difficulty' (0) to 'unable to do' (10). The function section has 11 questions regarding performance of specific activities and 4 questions on performing usual activities using the affected side. All the section scores are computed to obtain a global score out of 100. The higher the PREE total score the greater the pain and disability.

American Shoulder and Elbow Surgeons-Elbow Form: The American Shoulder and Elbow Surgeons-Elbow form (ASES-e) is a standardized elbow evaluation system that was developed by the Research Committee of the American Shoulder and Elbow Surgeons (ASES).[19] It has two parts: a physician form (cASES-e) and a patient-reported form (pASES-e). The Physician form has four components: motion, stability, strength and physical findings. The patient-reported form has three sections: pain, function and satisfaction. The pain section contains 5 questions which are rated using a 0-10 numeric rating scales where 0 is "no pain" and 10 is "worst pain ever". The function section contains 12 questions relating to function. The responses are scored on a four-point ordinal scale for both the affected and the unaffected arm, where, 0- unable to do; 1- very difficult to do; 2-somewhat difficult; 3-not difficult. The maximum score for the function of each arm is 36. A low score indicates worse function. The third section has 1 question to report patient-satisfaction with the surgery on a scale from 0 to 10. Subscales are not added to provide a total score.

Disabilities of the Arm, Shoulder and Hand Questionnaire: The Disabilities of Arm, Shoulder and Hand (DASH) Questionnaire is a 30 item patient-report questionnaire that evaluates impairments, activity limitations, as well as participation restrictions in both leisure activities and work due to disorders of the upper limb regardless of which arm is affected. [14] There are optional components that can be added to assess function with respect to sports and performing arts or work. The total DASH score ranges from 0 (no disability) to 100 (severe disability); with higher DASH scores indicating greater disability.

The Short Form 36-item Health Survey (SF-36): This is a 36 item questionnaire; yielding an eight subscale profile of scores; as well as a physical component summary score and a mental component summary score.[40] The SF-36 is a generic health status measure that has been used to evaluate quality of life in a wide spectrum of health conditions. It has more than a 1000 citations and has been translated in many languages. It is a valid and reliable health status instrument.[41] Higher scores represent better health (less disability).

Statistical Analysis

Data entry, quality checking and analysis (internal consistency, validity and sensitivity to change) were performed using the SPSS software version 19. The level of statistical significance was set at $p < 0.05$.

Reliability: Reliability is a generic term used to indicate both the homogeneity (internal consistency) of a scale and the reproducibility or stability (test–retest reliability) of scores.[6] The test-retest reliability for all three questionnaires was reported. In the PREE developmental study;[27] but internal consistency was not reported. Hence, only the internal consistency of these questionnaires was evaluated in this study. Cronbach's alpha was used to assess internal consistency of the PREE, the pASES-e and the DASH. 95% confidence intervals were also calculated. [7] The baseline scores were used to calculate internal consistency. Internal consistency is considered acceptable when Cronbach's alpha exceeds 0.70.[38]

Concurrent construct validity: To analyze Concurrent construct validity it was examined whether the PROs demonstrated expected relationships with other measures at a single point in time using Pearson's correlation coefficient (*r).* Pearson's correlation coefficients were

interpreted as follows: 0.00 to 0.19 = very weak correlation; 0.20 to 0.39 = weak correlation; 0.40 to 0.69 = moderate correlation; 0.70 to 0.89 = strong correlation; and 0.90 to 1 = very strong correlation.[10] Constructed hypotheses about the nature of the relationship between different questionnaires/subscales were prespecified as: strong correlations were expected between the most similar elbow questionnaires (the pASES-e and the PREE total and the pain scales); moderate to high correlations were expected between elbow questionnaires (the pASES-e and the PREE) and the DASH; low to moderate correlations were expected between elbow/DASH questionnaires and physical component summary scores of the SF-36; low correlations were expected between elbow/DASH questionnaires and mental component summary scores of the SF-36.

Longitudinal validity: Longitudinal validity is the extent to which changes on one measure will correlate with changes on another measure.[15] Longitudinal validity was evaluated by obtaining correlations between the baseline and 6 months change scores.[13] Pearson correlation coefficient was used and the correlations were interpreted in a manner similar to that of the concurrent construct validity metric above.

Sensitivity to change: Sensitivity to change is the ability of an instrument to measure real change in a clinical state.[23, 37] There is a lack of agreement on the best statistical method to analyze sensitivity to change. We used two distribution-based methods, the effect sizes (ES) and the standardized response means (SRM). The ES was calculated by dividing the mean change scores of the PREE, the pASES-e and the DASH in patients clinically identified as improved, by the standard deviation of their baseline scores.[18] In our study, included patients were clinically identified as improved or not improved by the attending surgeon during the follow up visit based on change in clinical signs and symptoms . The SRM was calculated by dividing the mean change scores of the PREE, the pASES-e and the DASH by the standard deviation of their change scores.[22] In addition, 95% confidence intervals around the point estimates were also calculated; and used to determine differences across sensitivity indices.[2] The size of the ES and SRM scores were interpreted as follows: 0.2 = small, 0.5 = moderate, 0.8 or higher = large.[3, 22]

Factor analysis: A principal components exploratory analysis with a varimax rotation was performed to investigate the factor structure of the PREE, the pASES-e and the DASH. Kaiser

criterion with Eigen values > 1, examination of the Scree plot as well as clinical interpretability were criteria applied to determine the number of components to be retained.[8] An item was considered to load on a given factor if the factor loadings were 0.4 or greater and was less than 0.4 for the other factors. Factor loadings over ±0.5 were considered as strong. Items that correlated more than 0.40 on more than one factor, without a differential of 0.2 were considered to be "cross loaded". The quality of the factor analysis was assessed using the Kaiser-Meyer-Olkin (KMO) test and Bartlett's test of sphericity. KMO tests the sampling adequacy; a high value (> 0.7) indicates that factor analysis is appropriate. Bartlett's test of sphericity tests the correlations between items. A significant value was considered an indication that our data were appropriate and the sampling was of high quality for factor analysis to be performed.[5]

RESULTS

Internal Consistency

The PREE, the pASES-e and the DASH exhibited excellent internal consistency with Cronbach's alpha values of 0.95 (95% C.I. 0.94- 0.96), 0.93 (95% C.I. 0.91-0.94) and 0.97 (95% C.I. 0.96-0.98) respectively.

Concurrent Construct Validity

In terms of concurrent construct validity, the strongest correlations were observed between the PREE total score, the PREE pain score, the PREE function score, the pASES-e pain score and the DASH total score (r = 0.91 to 0.72) (See Table.2). The PREE total score exhibited strong correlations with the pain scores of the pASES-e and the PREE (r = 0.87 and r = 0.82 respectively). The DASH (r = 0.54 and r = 0.54) and the PREE function scale (r = 0.45 and r = 0.49) correlated moderately with pain subscales of the pASES-e and the PREE. The function scale of pASES-e correlated the least with the pain scores of the pASES-e (r = -0.36) and the PREE (r = -0.33). Low to moderate inverse correlations were observed between the physical component summary score of the SF-36 and the three PROs (r = -0.46 to -0.38). The mental component summary score demonstrated a weak inverse relationship with all the three upper extremity PROs (r = -0.23 to -0.33) while its correlation with the pain scores were not

significant. All the correlations followed the expected relationships that were constructed prior to data analysis.

Longitudinal validity

Strong correlations (r >0.70) (See Table.3) were observed between change scores obtained from the pASES-e pain score and changes on the PREE score; weak correlations (r <0.40) were observed between changes on the pASES-e function score and all other change scores (r = 0.23 to 0.40). Moderate correlations (r = 0.41 to 0.69) were observed between the DASH change scores and the change scores of other questionnaires except with pASES-e function change score (r=0.23).

Sensitivity to change

All three patient-report questionnaires demonstrated large effect sizes ranging between 1.3 and 1.7. (See Table. 4) The pASES-e function scale had an ES of 1.3 (95% C.I. 1.1-1.5). Large SRMs were also obtained. The PREE [SRM=1.6 (95% C.I. 1.4-1.8)] and the DASH were similar [SRM=1.6 (95% C.I. 1.5-1.8)] and the pASES-e was slightly less sensitive to change when compared to the other questionnaires (pASES-e pain scale SRM =1.2 (95% C.I. 1.0-1.4); pASES-e function scale SRM= 1.09 (95% C.I. 0.9-1.3)). The confidence intervals were less overlapping, which means that the measures are quite different from each other.

Factor analysis

Four main components were identified when the factor structure of the PREE was analyzed explaining about 77.2% of the variance of the questionnaire's total score. These factors supported that the pain and usual items separated into individual subscales; items within the specific function subscale were separated into two components reflecting light and heavy activities. (See table 5) Analysis of the factor structure of the pASES-e revealed four factors. These four factors put together explained 74.4% of the variance of the pASES-e's main scores. These factors supported the fact that the pain items grouped under one factor while the questions related to the affected and unaffected arm grouped separately under the other factors. (See table 6) The DASH loaded onto 5 factors with more number of cross loadings than the other two questionnaires. All these 5 factors put together explained about 71.3% of the variance of the total score. (See table 7)

DISCUSSION

This study provides support for the internal consistency, concurrent construct validity, longitudinal validity, sensitivity to change and factor structure of two elbow specific questionnaires, the PREE and the pASES-e. It also provides further validation on the DASH, for use with elbow pathologies. The factor structure analysis, despite some inherent limitations in our analysis and sample size for this questionnaire, also contributes further evidence on the existing controversy of the factor structure of the DASH whether it is unidimensional or multidimensional.[11, 21, 29]

In the current study, Cronbach's alpha for the PREE was 0.95 with very narrow CI, which indicates an excellent internal consistency. This value is in line with that of the German version of the PREE which was 0.96.[16] Cronbach's alpha for the pASES-e was 0.93 in this study, which is comparable with the Cronbach's alpha value of 0.90 for the German version of the pASES-e.[17] Although the internal consistency of the DASH has not been reported specifically for elbow conditions; the high internal consistency and narrower CI found in this study is consistent with the Cronbach's alpha values obtained for the DASH in other studies focussing on shoulder and distal upper limb conditions.[12, 20]

We prespecified constructs around expected relationships between the questionnaires and subscales in this study similar to those reported in the PREE developmental study[27] and found similar findings. A new finding in this data is the fact that the functional subscale of the pASES-e was less correlated to other questionnaires. It is noteworthy that this subscale appeared to demonstrate poorer performance than the functional item subscale of the PREE since both are elbow specific questionnaires. We expect that the differences on content and measurement metrics in the numeric metric (0-10 vs 0-3) may have contributed to these findings. The pain subscale of the PREE and the pASES-e should correlate highly because they contain similar items (4 of 5 or 80% of the questions address the same content) and a similar 0-10 rating scale. Conversely, the PREE and pASES-e function scales are more different; around 50% of the items address the same content and the measurement metrics differ. These differences may have influenced the correlations obtained and the lower sensitivity to change scores evident in the pASES-e.

A notable difference between the pASES-e and other questionnaires is that there is no clear method for calculating a total score. It can be argued whether it is advisable to combine pain and disability subscales into a total score since these reflect two separate, although often correlated, constructs. However, the reporting of a single total score for pain and disability is common in studies of musculoskeletal conditions. If a total score were required it would make the most sense to equally rate pain and disability subscales of the pASES-e as this is consistent with the approach on the PREE; and other scales for the wrist[26] or shoulder.[31] It is not advisable to incorporate patient satisfaction scores into pain and disability questionnaires given the constructs are too diverse.[32]

There are no published reports describing the longitudinal validity of the PREE, pASES-e or the DASH in an elbow disorder population. Our findings of excellent correlation between the PREE change score, pASES-e pain change score and the DASH change score indicate that the change on these measures are strongly related. From a clinical perspective this suggests that people using different instruments would agree on whether patients had changed after treatment. The pASES-e function change score had the least correlation with other measures (r = 0.23 to 0.40). We expect that measurement scale difference also contributed to these lower correlations of the pASES-e. The ES and SRM demonstrated that all three questionnaires were highly sensitive to change. The 95% CI were less overlapping indicating that the measures are quite different. The CIs were also narrower, which can be attributed to the sample size. The pASES-e function scale had the lowest ES and SRM. The results of the current study confirm previous concern[27] that the pASES-e function subscale would be relatively less sensitive to change compared to other questionnaires measuring the same construct. Overall our findings suggest that the scoring metric of subscale of the pASES-e the function is suboptimal.

The factor analysed the principal components identified in all three questionnaires separately explained more than 70% of the variance of their total scores. The DASH had more items that cross loaded in comparison to the elbow scales. It is possible that DASH items are more generic and this may have contributed to the cross loading. However, five factors have being identified in this exploratory analysis- this undoubtedly also contributed to the higher level of cross loading. The suggestion that the DASH items are more generic was supported by Angst

et al[1] who reported that the DASH loaded more with the SF-36 than with the PREE or the pASES-e which provides futher verification than the DASH is more generalized than the PREE or ASES-e. The specificity of the pASES-e and of the PREE is consistent with these measures being joint-specific; and the reason some choose an elbow specific measures since the relevancy of the items might be specific to the concerns of patients with these disorders.

Previous literature has suggested that there is a theoretical relationship that exists between Cronbach's alpha and factor structure. Alpha is described as a function of the parameters of the hierarchical factor analysis model which allows for a general factor that is common to all of the items of a measure in addition to group factors that are common to some but not all of the items of a measure.[42] Cortina has suggested that high internal consistency could mean that the scale is unidimensional.[4] Our exploratory factor analysis found multiple factors for the DASH, which questions its unidimensionality. However, Cronbach's alpha was high indicating high internal consistency.

There has been debate in the literature about whether the DASH is unidimensional in structure.[11, 21, 29] Our study also found items of the DASH to load onto multiple factors indicating multidimensionality. We used an exploratory factor analysis because there have been no previous reports of factor analysis on the elbow questionnaires and on the DASH in an elbow sample; and hence no evidence to support a confirmatory factor analysis. More importantly, we wanted to examine DASH metrics using the same analysis as the PREE and ASES. If our primary purpose had been to evaluate the factor structure of the DASH we would have elected to use confirmatory factor analysis in a larger sample. Since the DASH has 30 items a sample size of 300 fulfills "the rule of thumb" of 10 subjects per item to provide a more thorough factor analysis. However, it is noteworthy that our KMO values (0.85 - 0.92) justified the adequacy of our sample size. Given our findings, a confirmatory factor analysis of DASH scores in a larger cohort of patients with elbow disorders is warranted to provide more definitive analysis of the factor structure of the DASH in this population.

In our study we observed that the pASES-e function sub-scale was least valid and sensitive to change. We suggest a few amendments to the pASES-e that might improve its performance in clinical research studies and make it comparable with similar questionnaires

reported in the literature. Firstly, we agree that pain and disability subscales should be separately reported in clinical research studies since these are different constructs. However, there are often reasons to also report a total score as a primary outcome measure of overall effect. When computing a total score, we would recommend equal weighting of the pain and disability subscales of the pASES-e to provide a total score comprised of 50% pain and 50% disability consistent with the approach used on the PREE, and other joint specific measures used in the upper extremity.[26, 30] We suggest that the patient satisfaction item should be separately reported; and not incorporated into a global score. We know that satisfaction is not related to other outcomes[32] in shoulder disorders; and may be more reflective of process than outcomes. Our second recommendation concerns the scaling for the function subscale since it appears to present some measurement limitations. A potential consideration is to change the scale to a 0-10 scale to make it the same metric as the pain subscale and other measures. This may also enhance sensitivity to change. We anticipate these changes would increase the utility of the pASES-e, and make it more correlated with the PREE.

The strengths of the current study include prospective evaluation of patients with elbow disorders at specified time points; the sample size is relatively large in comparison to other psychometric data reported in the literature and our KMO statistics indicate that we were adequately powered for factor analysis of the elbow questionnaires. Our study also has limitations. Only three subgroups of elbow patients were included, making it uncertain whether these results can be generalized to other elbow disorders. We did not specifically study test-retest reliability; since this has previously been reported to be high for all the 3 measures[27] and we cannot report whether this varied across the subgroups. The most important limitation of our study was that we did not administer a Global Rating of Change scale (GRC) to our patients. We used a distribution-based approach to calculate the sensitivity to change and were not able to calculate the clinically important difference (CID). Future studies should focus on including a wide variety of patients with different elbow pathologies. This would help improve the strength of the already available evidence to support the external validity of these questionnaires.

CONCLUSION

The results of this study have generated evidence to support the validity and sensitivity to change of the PREE, the pASES-e and the DASH. These results support the use of these three questionnaires by clinicians based on their needs and resources available. Minor changes to the scoring of the pASES-e should be investigated as a means to enhance its measurement properties.

Key Points

Findings

The PREE, the pASES-e and the DASH demonstrated strong construct validity and sensitivity to change in patients with elbow pathologies. The PREE and pASES-e demonstrated appropriate structural validity.

Implication

Clinicians may choose from these 3 scales to assess the status of patients with elbow disorders, and treatment results in change.

Caution

Only three subgroups of elbow patients were included, making it uncertain whether these results can be generalized to other elbow disorders. Clinically Important Difference could not be calculated because the GRC was not administered.

REFERENCES

1. Angst F, John M, Pap G, et al. Comprehensive assessment of clinical outcome and quality of life after total elbow arthroplasty. Arthritis Rheum. 2005;53(1):73-82 10.1002/art.20911.

2. Beaton D, Richards RR. Assessing the reliability and responsiveness of 5 shoulder questionnaires. J Shoulder Elbow Surg. 1998;7(6):565-72.

3. Cohen J. *Statistical power analysis for the behavioral sciences.* 2nd ed. Hillsdale, N.J.: L. Erlbaum Associates; 1988.

4. Cortina JM. What is coefficient alpha? an examination of theory and applications. J Appl Psychol. 1993;78(1):98-104 10.1037/0021-9010.78.1.98.

5. *Cudeck R.* Exploratory factor analysis. In: *Handbook of multivariate experimental psychology.* Elsevier Science publishers BV; 2000. p. 265-96.

6. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. statistics and strategies for evaluation. Control Clin Trials. 1991;12(4 Suppl):142S-58S.

7. Duhachek A, Lacobucci D. Alpha's standard error (ASE): An accurate and precise confidence interval estimate. J Appl Psychol. 2004;89(5):792-808 10.1037/0021-9010.89.5.792.

8. Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. Psychol Assess. 1995;7(3):286-99.

9. Food and Drug Administration.Patient reported outcome measures: Use in medical product development to support labeling claims. 2009; (11/8/2011).

10. Fowler J, Chevannes M, Jarvis P. *Practical statistics for nursing and health care.* New York; Chichester: Wiley; 2002.

11. Franchignoni F, Giordano A, Sartorio F, Vercelli S, Pascariello B, Ferriero G. Suggestions for refinement of the Disabilities of the Arm, Shoulder and Hand outcome measure (DASH): A factor analysis and Rasch validation study. Arch Phys Med Rehabil. 2010;91(9):1370-7 10.1016/j.apmr.2010.06.022.

12. Gummesson C, Atroshi I, Ekdahl C. The Disabilities of the Arm, Shoulder and Hand (DASH) outcome questionnaire: Longitudinal construct validity and measuring self-rated health change after surgery. BMC Musculoskelet Disord. 2003;4:11 10.1186/1471-2474-4-11.

13. Guyatt G, Walter S, Norman G. Measuring change over time: Assessing the usefulness of evaluative instruments. J Chronic Dis. 1987;40(2):171-8.

14. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: The DASH (Disabilities of the Arm, Shoulder and Hand) [corrected]. the Upper Extremity Collaborative Group (UECG). Am J Ind Med. 1996;29(6):602-8 2-L.

15. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: A critical review and recommendations. J Clin Epidemiol. 2000;53(5):459 http://www.sciencedirect.com/science/article/pii/S0895435699002061. 10.1016/S0895-4356(99)00206-1".

16. John M, Angst F, Awiszus F, King GJ, MacDermid JC, Simmen BR. The American Shoulder and Elbow Surgeons elbow questionnaire: Cross-cultural adaptation into german and evaluation of its psychometric properties. J Hand Ther. 2010;23(3):301,13; 314 10.1016/j.jht.2010.03.001.

17. John M, Angst F, Pap G, Junge A, Mannion AF. Cross-cultural adaptation, reliability and validity of the Patient Rated Elbow Evaluation (PREE) for German-speaking patients. Clin Exp Rheumatol. 2007;25(2):195-205.

18. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care. 1989;27(3 Suppl):S178-89.

19. King GJ, Richards RR, Zuckerman JD, et al. A standardized method for assessment of elbow function. research committee, American Shoulder and Elbow Surgeons. J Shoulder Elbow Surg. 1999;8(4):351-4.

20. Kitis A, Celik E, Aslan UB, Zencir M. DASH questionnaire for the analysis of musculoskeletal symptoms in industry workers: A validity and reliability study. Appl Ergon. 2009;40(2):251-5 10.1016/j.apergo.2008.04.005.

21. Lehman LA, Woodbury M, Velozo CA. Examination of the factor structure of the Disabilities of the Arm, Shoulder, and Hand questionnaire. The American Journal of Occupational Therapy. 2011;65(2):169-78 10.5014/ajot.2011.000794.

22. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care. 1990;28(7):632-42.

23. Liang MH. Longitudinal construct validity: Establishment of clinical meaning in patient evaluative instruments. Med Care. 2000;38(9, Supplement II: Health Outcomes Methodology: Symposium Proceedings):pp. II84-II90 http://www.jstor.org.proxy1.lib.uwo.ca:2048/stable/3768066.

24. L'Insalata JC, Warren RF, Cohen SB, Altchek DW, Peterson MG. A self-administered questionnaire for assessment of symptoms and function of the shoulder. J Bone Joint Surg Am. 1997;79(5):738-48.

25. Patient-rated elbow evaluation [homepage on the Internet]. Available from: http://www.srs-mcmaster.ca/Portals/20/pdf/research_resources/PREE.pdf.

26. MacDermid JC. Development of a scale for patient rating of wrist pain and disability. J Hand Ther. 1996;9(2):178-83.

27. MacDermid JC. Outcome evaluation in patients with elbow pathology: Issues in instrument development and evaluation. J Hand Ther. 2001;14(2):105-14.

28. MacDermid JC, Roth JH, McMurtry R. Predictors of time lost from work following a distal radius fracture. J Occup Rehabil. 2007;17(1):47-62 10.1007/s10926-007-9069-0.

29. MacDermid JC, Wessel J, Humphrey R, Ross D, Roth JH. Validity of self-report measures of pain and disability for persons who have undergone arthroplasty for osteoarthritis of the carpometacarpal joint of the hand. Osteoarthritis Cartilage. 2007;15(5):524-30 10.1016/j.joca.2006.10.018.

30. Overend TJ, Wuori-Fearn JL, Kramer JF, MacDermid JC. Reliability of a patient-rated forearm evaluation questionnaire for patients with lateral epicondylitis. Journal of Hand Therapy. 1999;12(1):31-7 10.1016/S0894-1130(99)80031-3.

31. Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. Arthritis Care Res. 1991;4(4):143-9.

32. Roy JS, Macdermid JC, Goel D, Faber KJ, Athwal GS, Drosdowech DS. What is a successful outcome following reverse total shoulder arthroplasty? Open Orthop J. 2010;4:157-63 10.2174/1874325001004010157.

33. Sathyamoorthy P, Kemp GJ, Rawal A, Rayner V, Frostick SP. Development and validation of an elbow score. Rheumatology (Oxford). 2004;43(11):1434-40 10.1093/rheumatology/keh367.

34. Slevin ML, Plant H, Lynch D, Drinkwater J, Gregory WM. Who should measure quality of life, the doctor or the patient? Br J Cancer. 1988;57(1):109-12.

35. Solway S, Beaton DE, McConnell S, Bombardier C. The DASH outcome measure user's manual(2nd ed.). Toronto: Institute for Work & Health. 2002.

36. Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: A review. J Clin Epidemiol. 1992;45(7):743-60.

37. Stratford PW, Binkley JM, Riddle DL. Health status measures: Strategies and analytic methods for assessing change scores. Phys Ther. 1996;76(10):1109-23.

38. Streiner DL, Norman GR. *Health measurement scales :A practical guide to their development and use.* 4th ed. Oxford: Oxford University Press; 2008.

39. Turchin DC, Beaton DE, Richards RR. Validity of observer-based aggregate scoring systems as descriptors of elbow pain, function, and disability. J Bone Joint Surg Am. 1998;80(2):154-62.

40. *Ware JE, Snow KK, Kosinski M, Gandek B, eds.* SF-36 health survey manual and interpretation guide. Boston (MA): The Health Institute, New England Medical Center; 1993.

41. Ware JE,Jr. SF-36 health survey update. Spine (Phila Pa 1976). 2000;25(24):3130-9.

42. Zinbarg RE, Revelle W, Yovel I, Li W. Cronbach's α, Revelle's β, and Mcdonald's ω*H*: Their relations with each other and two alternative conceptualizations of reliability. Psychometrika. 2005;70(1):123-33 http://dx.doi.org/10.1007/s11336-003-0974-7.

Table 1: Baseline patient characteristics

Characteristic	Values
Subjects (n)	128
Male/Female (n)	84:44
Age in years*	46.53(12.77)
Age range	21 –79 years
Dominant hand (R:L)	118:10
Injured elbow (R:L)	57:66
Surgical procedure	**n (%)**
Biceps tendon repair	62(48.44)
Radial head arthroplasty	36(28.13)
Radial head fixation	30(23.44)
WCIB case	**n (%)**
Yes	24 (18.75)
No	101 (78.90)
Pending	3 (0.02)

* The values are expressed as mean and standard deviation (SD). WCIB- Worker's Compensation Insurance Board (Canada)

Table 2: Concurrent construct validity of the PREE, the pASES-e and the DASH

Pearson's correlation coefficient (*r*) (N=128)	**PREE Pain**	**PREE Function**	**PREE Total**	**pASES-e Pain**	**pASES-e function affected**	**DASH Total**
PREE Pain	1					
PREE Function	0.49	1				
PREE total	0.82	0.91	1			
pASES-e pain	0.87	0.45	0.73	1		
pASES-e function affected	-0.36	-0.68	-0.62	-0.33	1	
DASH total	-0.54	0.72	0.75	0.54	-0.55	1
SF36 Physical component summary	-0.21**	-0.42	-0.38	-0.27	0.38	-0.46
SF36 Mental component summary	-0.18*	-0.21**	-0.23**	-0.14*	0.24**	-0.33

All correlations are significant to the level of $p<0.01$ unless indicated * Not significant; **$p<0.05$; PREE- Patient rated elbow evaluation; pASES-e American society of elbow surgeons-Elbow form; DASH- Disabilities of the arm, shoulder and hand; SF-36-MCS- Short form-36 Mental component summary score; SF-36-PCS- Short form-36 Physical component summary score.

Table 3: Longitudinal validity of the PREE, the ASES-e and the DASH

Pearson's correlation coefficient (r) (N=128)	PREE Pain Change Score	PREE Function Change Score	PREE total Change score	pASES-e Pain change score	pASES-e function change score	DASH change score
PREE Pain Change score	1					
PREE Function Change Score	0.40	1				
PREE total change score	0.81	0.86	1			
pASES-e pain change score	0.85	0.41	0.74	1		
pASES-e function change score	0.24	0.40	0.33	0.24	1	
DASH change score	0.41	0.62	0.62	0.46	0.23	1

All correlations are significant to the level of $p<0.01$. Change score- Difference between baseline and six months scores; PREE- Patient rated elbow evaluation; pASES-e American society of elbow surgeons- Elbow form; DASH- Disabilities of the arm, shoulder and hand.

Table 4: Effect sizes and Standardized response mean for the three questionnaires

Outcome measure	Baseline score*	6 months score*	Change score*	ES (95% C.I.)	SRM (95% C.I.)
PREE Pain	25.13 (11.42)	9.43 (9.99)	15.86 (12.67)	1.4 (1.2 - 1.6)	1.3 (1.0 - 1.4)
PREE Function	27.99 (15.53)	5.93 (11.16)	22.31 (14.57)	1.4 (1.3 - 1.6)	1.5 (1.4 - 1.7)
PREE Total	53.12 (23.39)	15.36 (19.72)	38.75 (24.34)	1.7 (1.5 - 1.8)	1.6 (1.4 - 1.8)
pASESe-Pain	24.24 (11.36)	8.31 (9.53)	16.02 (13.21)	1.4 (1.2 - 1.6)	1.2 (1.0 - 1.4)
pASESe-Function	14.68 (11.33)	28.76 (9.08)	14.61 (13.45)	1.3 (1.1 - 1.5)	1.1 (0.9 - 1.3)
DASH	46.33 (23.32)	13.2 (15.47)	33.29 (20.34)	1.4 (1.3 - 1.6)	1.6 (1.5 - 1.8)

*The values are expressed as mean and standard deviation (SD); PREE- Patient rated elbow evaluation; pASES-e American society of elbow surgeons- Elbow form; DASH- Disabilities of the arm, shoulder and hand; ES Effect Size; SRM Standardized Response Means; 95% C.I. – 95% Confidence Interval

Table 5: Summary of exploratory factor analysis of the PREE using Principal Component Analysis with Varimax rotation

Item (R^2=77.2%)	1	2	3	4
Comb my hair	0.83			
Eat with a fork or spoon	0.82			
Use my arm to rise from a chair	0.62			
Use a telephone	0.85			
Do up buttons on the front of my shirt	0.87			
Wash my opposite armpit	0.79			
Tie my shoe	0.75			
Turn the doorknob and open a door	**0.67**			**0.59**
Pain - When it is at its worst		0.81		
Pain - At rest		0.74		
Pain - When lifting a heavy object		0.68		
Pain - When doing a task with repeated elbow movement		0.75		
How often do you have pain?		0.77		
Personal activities (dressing, washing)			0.71	
Household work (cleaning, maintenance)			0.81	
Work (your job or everyday work)			0.85	
Recreational activities			0.85	
Pull a heavy object				0.78
Carry a 10lb object with my arm at my side				0.59
Throw a small object, such as a tennis ball				0.66

*Cross loaded items are bolded; PREE- Patient rated elbow evaluation; R^2 – percentage of variance explained

Table 6: Summary of exploratory factor analysis of the pASES-e using Principal Component Analysis with Varimax rotation

Item (R^2 = 74.4%)	1	2	3	4
Do up top button on shirt (Affected)	0.91			
Manage toileting (Affected)	0.87			
Comb hair (Affected)	0.90			
Tie shoes (Affected)	0.91			
Eat with utensil (Affected)	0.88			
Carry a heavy object (Affected)	0.69			
Rise from chair pushing with arm (Affected)	0.87			
Do usual sport (Affected)	0.65			
Do usual work (Affected)		0.88		
Do up top button on shirt (Unaffected)		0.89		
Manage toileting (Unaffected)		0.87		
Comb hair (Unaffected)		0.88		
Tie shoes (Unaffected)		**0.58**		**0.47**
Eat with utensil (Unaffected)		0.93		
Carry a heavy object (Unaffected)		0.76		
Rise from chair pushing with arm (Unaffected)		0.88		
Pain - When it is at its worst			0.85	
Pain - At rest			0.71	
Pain - When lifting a heavy object			0.82	
Pain - When doing a task with repeated elbow movement			0.84	
Pain- At night			0.72	
Do usual work - describe (unaffected)		**0.49**		**0.76**
Do usual sport – describe (Unaffected)				0.86

*Cross loaded items are bolded; pASES-e American society of elbow surgeons- Elbow form; R^2 – percentage of variance explained

Table 7: Summary of exploratory factor analysis of the DASH using using Principal Component Analysis with Varimax rotation

Item (r^2=71.3%)	1	2	3	4	5
Write.	0.45	0.42			
Prepare a meal.	0.59				
Make a bed.	**0.66**		**0.45**		
Wash or blow dry your hair.	0.68	0.49			
Wash your back.	0.61				
Put on a pullover sweater	0.69				
Use a knife to cut food	0.69				
Recreational activities which require little effort	0.69				
Manage transportation needs	0.62				
Sexual activities.	0.71				
Stiffness in your arm, shoulder or hand.	**0.60**			**0.54**	
Turn a key.		0.74			
Push open a heavy door.		0.72			
Place an object on a shelf above your head.		0.76			
Carry a shopping bag or briefcase.		0.85			
Carry a heavy object (over 10 lbs).		0.81			
Change a lightbulb overhead.		0.70			
Do heavy household chores	**0.53**		**0.60**		
Garden or do yard work.	**0.52**		**0.62**		
Open a tight or new jar.			0.63		
Recreational activities in which you take some force or impact through your arm, shoulder or hand.			0.73		
Recreational activities in which you move your arm freely			0.73		
Arm, shoulder or hand pain.				0.78	
Arm, shoulder or hand pain when you performed any specific activity.				0.76	
Tingling (pins and needles) in your arm, shoulder or hand.				0.45	
Weakness in your arm, shoulder or hand.				0.73	
During the past week, were you limited in your work or other regular daily activities as a result of your arm, shoulder or hand problem?				**0.46**	**0.52**
During the past week, *to what extent* has your arm, shoulder or hand problem interfered with your normal social activities with family, friends, neighbours or groups?					0.81
During the past week, how much difficulty have you had sleeping because of the pain in your arm, shoulder or hand?					0.65
I feel less capable, less confident or less useful because of my arm, shoulder or hand problem.					0.63

*Cross loaded items are bolded; DASH- Disabilities of the arm, shoulder and hand; R^2 – percentage of variance explained

CHAPTER 3: MEASUREMENT OF ELBOW PAIN AND FUNCTION: A SYSTEMATIC REVIEW OF THE PSYCHOMETRIC PROPERTIES OF TWO ELBOW SELF-REPORT MEASURES[1]

ABSTRACT

Objective: Patient Rated Elbow Evaluation (PREE) and the self-report section of the American Shoulder Elbow Surgeons – elbow form (pASES-e) are two patient reported outcome measures (PROM) commonly used to assess pain and disability arising from elbow disorders including degenerative and rheumatoid arthritis. The objective of this current study is to systematically review and summarize the quality and the content of the evidence that is available on their psychometric properties.

Methods: We systematically searched the online databases PubMed, EMBASE, ProQuest, CINHAL, Up-to-date, Dissertations and thesis and Google Scholar. 91 articles were retrieved and after screening, 9 were included in the final analysis. Data extraction and quality appraisal was performed by two independent raters. Descriptive synthesis of the reviewed studies was done.

Results: 7 of the 9 studies had a quality score of 75% or higher. Agreement between the raters was good (Kappa 0.81). Both the PROMs did not demonstrate any floor and ceiling effects except for the satisfaction sub scale of the pASES-e. Factor analysis revealed multi-dimensionality in the function sub scale for both the PROs. Construct validity was good with correlations above 0.70. Both were highly reliable with ICCs > 0.90. They were also highly responsive with an effect size (ES) and standardized response mean (SRM) above 1. The Minimal Clinical Important Difference (MCID) were not estimated for both measures.

Conclusion: The PREE and the pASES-e have been established to be valid, reliable and sensitive to change in both clinical and research settings based on high quality evidence.

A version of this work has been submitted for journal publication: **Vincent, JI**, MacDermid, JC, King GJW, Grewal, R. Measurement of elbow Pain and Function: a systematic review of the psychometric properties of two elbow self-report measures. Arthritis Care & Research. (2014)

SIGNIFICANCE AND INNOVATIONS

- Thorough systematic review on the psychometric properties of the two elbow self-report measures.
- Conclusions are based on high quality evidence.
- Both the PREE and the pASES-e have demonstrated optimal levels of psychometric properties supporting their continued use in both clinical and research settings.

INTRODUCTION

Patient reported outcome measures (PROM) are commonly used to measure pain and disability, which is in line with the patient centered care where patients are empowered to make informed decisions regarding their health status and its management.(1, 2) PROMs should provide a valid and reliable estimate of a patient's status, and should be able to detect change in a patient's state of health as it evolves. Hence, these measures require rigorous evaluation of their measurement properties. It is important for clinicians to ensure that the PROMs they use are valid and reliable, so that they can be confident about the results they obtain and the inferences they make.(3) Utility is another issue to consider with the use of PROMs. Administrative burden (i.e. time, cost and ease of calculating final scores) and patient burden (i.e. reading level, ease of completion, time required) are important factors that determine the utility of a PROM.

Evidence about the measurement properties of requires testing of different properties, patient populations and contexts so which is rarely accomplished in a single study. Therefore, it is important to identify, evaluate and synthesize the pool of evidence that informs our understanding in measurement performance of different tools. Systematic reviews of clinical measurement studies should provide an unbiased synthesis of the body of literature supporting specific applications of different PROMs and thereby assist clinicians by accessing the best evidence for selection and application of PROMs for their practice.

The Patient Rated Elbow Evaluation (PREE) (4, 5) and the self-report section of the American Shoulder Elbow Surgeons – elbow form (pASES-e).(6) are the two joint specific PROMs that are commonly used to assess the pain and disability from elbow disorders. These two measures were developed in the 1990's and are being used consistently by surgeons and hand therapists treating elbow disorders; and upper extremity researchers alike across the globe. Recently, these measures have both been recommended as a core set of measures to assess patients with elbow disorders.(7) Cross cultural adaptations of these tools are emerging.(8-10)

A previous systematic review has focused on evaluating all the rating systems of the elbow.(11) There were some potential pitfalls in that study. Firstly, they did not have a specific research question with specified inclusion and exclusion criteria. Further, they only searched one database (PubMed). This created the potential for selection bias and missing evidence that might

have been obtained through other electronic databases (i.e. Scopus, ProQuest etc.). Previous studies have shown that a search of PubMed alone does not identify all relevant articles.(12, 13) They did not include studies from languages other than English or versions of the PROMs in other languages. This may have led to language bias. Some of the articles that were missed in that review had the potential to change the conclusions on the quality of the PREE that was laid out in that review. The pASES-e was not included in the review. Hence the purpose of the current study was to perform a comprehensive systematic review of the literature and summarize the quality and the content of the evidence that is available on the psychometric properties of the PREE and the pASES-e.

MATERIALS AND METHODS

Description of the measures

Patient Rated Elbow Evaluation: The PREE was developed in the year 2000. It is a twenty item PROM which evaluates pain and disability arising from elbow disorders.(4, 5) It has 2 sections namely, pain and function. The pain section consists of five items and the function section has fifteen items. The function section is further sub-divided into specific activities (11 items) and usual activities sub scales (4 items). Each of the items are scored on a 0 – 10 scale. The total score is calculated out of 100, where pain and disability are equally weighted. The higher the PREE total score the greater the pain and disability.(14)

American shoulder and elbow surgeons-elbow form (self-report part) (pASES-e): The pASES-e is an eighteen item PRO with three sections– pain, function and satisfaction.(6) The pain section has five items scored on a scale from 0 to 10. The function section has 12 items scored on a 0-3 ordinal scale and evaluates both the affected and unaffected arm. The greater the total score, the greater the disability. The satisfaction section has 1 item asking patients to rate their satisfaction from treatment on a scale of 0 to 10. There is no total score for the pASES-e.

Search strategy: Electronic databases PubMed, EMBASE, ProQuest, CINHAL, Up-to-date; Dissertations and thesis, Google Scholar along with specific journals like Hand, JOSPT, Journal of Hand Therapy were searched; also reference lists were retrieved from articles and were used to identify potential articles that could have been missed during the regular search. Since these questionnaires were developed after 1995, these databases were searched from Jan 1995 to Apr

2014, in all languages. Three groups of search terms were used in various combinations using 'AND's and 'OR's. (See Appendix 1) The PRISMA guidelines (15) were followed in reporting the search strategy. (See Figure 1)

Selection criteria: Articles were included in the review if they met the following inclusion criteria: 1) measurement of at least one psychometric property of either the PREE or the ASES-e, 2) published in any language and 3) inclusion of subjects with any elbow disorders. In the first stage of the literature search, the title/abstracts from the search yield were screened by two independent reviewers based on the inclusion criteria. In the next stage, full manuscripts were retrieved for the selected articles. After this, the final set of articles that met the inclusion criteria were selected and subjected to quality appraisal.

Quality Appraisal: A pair of raters independently conducted a quality appraisal of the articles retrieved based on the inclusion criteria. The quality appraisal of the included studies was assessed using a previously developed quality appraisal tool developed by MacDermid- the Quality Appraisal for Clinical Measurement Studies (QA-CMS).(16) The QA-CMS has demonstrated high inter-rater reliability (ICC > 0.90).(17, 18) Once the quality was appraised, data was extracted using a data extraction form.(16) The reviewers met with the developer of the quality appraisal tool (JMD) and performed a calibration review. Every item of the appraisal tool was briefed to the reviewers to clarify the intended meaning and interpretation. Then the articles were independently appraised by the two reviewers. Once this process was complete, both the reviewers met for discussion to achieve consensus on individual items of appraisal. When consensus was not achieved, a third reviewer helped to resolve the disagreement. The studies were arranged in the descending order of quality scores. The kappa value was calculated to assess the overall inter-rater agreement before consensus was achieved.(19, 20) The inter-rater reliability of the quality assessment tool was measured using the interclass correlation co-efficient (ICC). (21-24)

RESULTS

Search yield: The search of the above mentioned sources yielded 91 articles. Once the duplicates and irrelevant articles were removed 32 articles remained for screening using the inclusion and exclusion criteria. Out of the thirty two articles that were screened, seventeen did

not meet the selection criteria, leaving fifteen articles for which full text was obtained. After reviewing the full text, six additional articles were excluded. Finally, 9 articles remained which were then included for quality appraisal. (See Figure 1) The PREE was studied more frequently than the pASES-e. Out of these nine articles 4 focussed specifically on the PREE; (7, 8, 10, 25, 26) only one was specific to pASES-e; (9) and the remaining four included both the PROMs.(5, 25, 27, 28) The synthesis of individual studies and the psychometric properties they are measuring are described in Table 1. All the studies included patients who underwent surgical intervention.

Quality rating: The quality score for the included articles ranged from 54% to 96%, with 78 % of articles having a score of 75% or higher. (See Table 2) The quality assessment tool exhibited excellent inter-rater reliability (ICC 0.94; 95% CI 0.91 – 0.96). The overall agreement between the raters was good (Kappa 0.81 $p<0.001$). The common design flaws observed with the reviewed articles were: most studies did not have a valid justification for sample size, studies did not explicitly report their hypothesis, and the studies lacked of clarity in reporting the type and sub-type of the psychometric properties being measured. For example, most of the studies mentioned that they were assessing validity and reliability but did not explicitly mention the type of validity or reliability they were intending to measure. Additionally, most studies did not report the estimates of measurement error like the SEM or use confidence intervals. As well, modern psychometric methods like item response, response shift or Rasch analysis were seldom used and conclusions were vague and non-specific.

Readability, language and cross cultural translation/adaptation: The PREE and the pASES-e both have been reported to be easy to read and understand.(28) (Table 3) The reading level for the PREE was equivalent to that of grade 6 (Flesch level 6.3).(29) With regards to missing responses to items, a minimal completion criterion of 65% was set for the PREE.(10) There was no specific pattern of missing responses to items.(8, 10) For the pASES-e one study reported items 18 and 19 (do usual work or sport) (27.5% to 33%) were missing.(9) The PREE has been translated into German (10) and Japanese (8) and has been reported to have exhibited psychometric properties that were similar to the English version. The German translation of the pASES-e has exhibited psychometric properties equivalent to that of the original English version.(9)

Administrative burden: Administrative burden was reported in only one study. Angst et al reported that it took only 3 minutes each to complete the PREE and the pASES-e.(28)

Floor and ceiling effects: The floor and ceiling effects of the two questionnaires have been reported. Angst et al(28) observed a ceiling of 15%, 4% and 4% for the pain subscale, function sub scale and the total PREE score respectively; while observing a ceiling of 14% for the pain subscale, 4% for function affected sub scale and 43% for satisfaction subscale. For the Japanese version of the PREE, no or very low (4%) floor and ceiling effects were observed.(8) While for the German version of the PREE, the ceiling ranged from 5.3% to 14.3%.(10)

Factorial validity: The PREE pain scale has been reported to be unidimensional both in the English version (25) and the Japanese version.(8) But the function sub scale of the English version loaded onto three factors while its Japanese counterpart loaded onto two factors.(8, 25) For the pASES-e the principle component analysis with varimax rotation revealed that the pASES-e loaded onto four different factors.(25)

Construct validity: Construct convergent validity has been established for the PREE and the pASES-e. Their strength of relationship with the DASH, Quick DASH and the SF-36 were consistently studied. For the PREE, high correlations ($r > 0.70$) were reported with the Quick DASH, DASH, pASES-e pain subscale, pASES-e total score, PREE pain sub scale and PREE function sub scale(5, 7, 8, 10, 25, 26, 28); moderate correlations ($r = 0.30 – 0.70$) have been established with the Clinician section of the ASES-e (cASES-e), pASES-e function sub scale and SF-36 physical component sub scale.(5, 10, 25, 28) The pASES-e total score correlated highly with the DASH; PREE function score and PREE total score (9) and moderately with the Clinician part of the ASES and SF-36 physical component sub scale.(5, 9, 28) Construct divergent validity for both the PROM has been established with the Mental Component Summary Score of the SF-36.(5, 7-10, 25, 28) The strength of correlations of individual sub scales with various constructs is shown in Table 3.

Reliability:

Internal consistency: The internal consistency of the PREE has been shown to be slightly higher than the pASES-e. For the PREE, Vincent et al (25) reported a Cronbach's alpha of 0.95 (95%CI 0.94 – 0.96), while Hanyu et al (8) and John et al (10) reported 0.97 and 0.96

respectively for the Japanese and the German version. For the pASES-e the Cronbach's alpha value was reported to be 0.93 (95% CI 0.91 – 0.94) (25) for the English version and 0.90 for the German version.(9) (See Table 4)

Test-retest reliability: Relative reliability of the PREE has been well established and has been shown to have excellent reliability co-efficients. For the English version, the ICCs of 0.95 (95% CI 0.86 – 0.98) for the total score and for the individual sub scales ICCs ranging from 0.88- 0.89 have been reported.(5) For the Japanese and the German version, the ICCs varied from 0.73 to 0.94(8, 10) for the different sub scales and the total score. (See Table 4) The pASES-e has also demonstrated acceptable levels of relative reliability. The ICCs for the various sub scales of the English version ranged from 0.64 to 0.90.(5) While for the German version higher ICCs were reported (0.74 to 0.93).(9) (See Table 4)

Minimal Detectable Change (MDC) and Standard Error of Measurement (SEM): Absolute reliability has been reported only for the German version of the pASES-e. (Table 4) The SEM for the sub scales ranged from 1.62 (95% CI -2.05 – 3.05) (satisfaction sub scale) to 11.62 (95% CI -23.74 – 23.28) (Function unaffected subscale) and 6.23 (95%CI -15.8 – 9.88).(9) $MDC_{95\%}$ for the pASES-e was 16.05 points.(9)

Responsiveness: Both the PREE and the ASES-e have been reported to be highly sensitive to change with effect sizes (ES) and standardized response means (SRM) above 0.80. The PREE was slightly higher when compared to the pASES-e. (Table 5) The ES and the SRM for the PREE, as reported in three studies, ranged from 1.7-1.1 and 1.3-1.6 respectively.(8, 25, 27) Angst et al(27) was the only study that has reported the ES and SRM for the pASES-e and they were 1.32 and 1.17 respectively.(27)

Only one study has established the longitudinal construct validity of the PREE and pASES-e.(25) The change scores of the two measures and the DASH were used in the analysis. (Table 5) For the PREE, the correlations ranged from 0.77 with the pASES-e pain scale to 0.33 with the pASES-e function scale and demonstrated a moderate correlation (r = 0.62) with the DASH. For the pASES-e, the pain sub scale exhibited high to moderate correlations with the PREE and the DASH (0.41 to 0.85). The function subscale correlated the least with correlations ranging from 0.23 to 0.40.

Angst et al has studied the sensitivity and specificity of the two measures.(27) The PREE and the pASES-e demonstrated almost similar sensitivity indices. (Table 5) The PREE discriminated between much better and the other categories with an AUC of 0.68 (95% CI 0.57 – 0.81) and it was 0.67 (95%CI 0.53 – 0.80) for the pASES-e which supports the power of discrimination of the measure. The PREE was more specific (0.71 Vs 0.69) and the pASES-e (0.65 Vs 0.63) was more sensitive.(27)

DISCUSSION

This study from its review of 9 articles concluded that strong clinical measurement properties exist for both the PREE and the pASES-e and that one cannot be recommended over the other. Despite the strong measurement properties reported we identified gaps in the current evidence for both the ASES-e and the PREE. The studies reviewed were of high quality which indicates that the recommendations that are generated from this review are based on quality clinical measurement research.

Content validity was reported as being acceptable for both PROMs as multiple studies suggested that 'all items were easy to understand'. However, the content of these PROMs have not been adequately analyzed using formal content analysis strategies. This is not unexpected since many PROM's have not used rigorous and structured content analysis during development or evaluation. It is recommended that the content of these questionnaires be analysed against the international standard set up by the World Health Organisation called the International Classification of Functioning Disability and Health (ICF). (30) Other methods that can provide a structured approach to evaluate content such as cognitive interviews, qualitative studies, concept mapping, etc. have not been reported as ways of analyzing content of either measure. Despite this lack of formal evaluation of content, face validity suggests that they both measure pain and disability in patients with elbow conditions.

Floor and ceiling effects can pose potential problems because when patients score at the extremes of a scale, further worsening or improvement may not be detected. A floor or ceiling effect has been defined as being problematic if more than 15% of the participants are at either ends of the spectrum; (31) however this was not a problem for the PREE and the pASES-e except for the satisfaction section for the pASES-e. It has been previously argued that

satisfaction is a concept that is less relevant to the constructs of pain and disability(32, 33) and should be considered as a separate section and efforts should not be made to include the score from the satisfaction section when total score calculation is attempted for pASES-e.(25)

Unidimensionality is another important clinical measurement property of outcome measures since measures are typically defined as measuring a specific construct and should be able to measure that as a single trait. (34) Exploratory factor analyses of both the PREE and the pASES-e suggested that they were not unidimensional. However, internal consistency was high (>0.90) for both the PREE and the pASES-e supporting unidimensionality. These findings need to be explored further using powerful statistical methods like confirmatory factor analysis, structural equation modeling or Rasch analysis.(35-38) The confirmatory analyses can determine the dimensionality of these measures and may suggest amendments that may render these measures unidimensional if they are not.(34, 37)

The PREE and the pASES-e have demonstrated expected correlations (>0.70) with other measures measuring a similar construct like the DASH, Quick DASH and between themselves. This indicates that these scales are measuring similar constructs and can be used solely based on the clinical need and ease of use. Both these measures have demonstrated moderate correlations with the physical component summary score of the SF-36, underlining the effect of disability on overall physical health status. There is little evidence in the literature surrounding the known group validity of these two measures making it difficult to make interpretations in specific populations. Future studies are needed to assess the validity of these measures on using additional clinical populations.

Both these measures have been deemed to be fit for use in either individual assessments or for use in a group as they have demonstrated high test-retest reliability ICC>0.90. However, these indices are measures of relative reliability and not absolute reliability. For a measure to be clinically useful, measures of absolute reliability such as the SEM and MDC should be provided.(21) These indices create estimates that are free of measurement error and capture changes that are important clinically and for the patients themselves. SEM and MDC have been calculated only for the German version of the pASES-e.(9) Another measure of reliability that was reported for both the PROMs is the Bland and Altman technique; it is used to assess reliability and can detect bias between assessments.(39) Future studies should focus on deriving

these measures to enhance clinical applicability by being able to detect and quantify change as it occurs.

The PREE and the pASES-e are sensitive to change established by large ES and SRM from multiple studies.(8, 25, 27) Therefore, these two measures can detect true change in the patent's status as it happens.(40) SRM and ES are distribution based methods and can only capture real change.(41, 42) To capture the clinically important change, future studies need to be conducted that use anchor based methods such as the Global Rating of Change (GRoC) to calculate minimal clinically important difference (MCID).(21, 40, 42) This is a vital characteristic of any measure that would help increase the tools clinical utility by increasing the confidence of the clinician or researcher when interpreting the results of an outcome measure.

A report on the quality assessment tools used to evaluate the quality of systematic reviews identified 86 different tools; it recommended that these tools should be developed using rigorous processes, based on evidence and they should be reliable.(43). A fundamental component of systematic reviews is that they assess study quality or risk of bias. The two most commonly used appraisal tools in clinical measurement research are the QA-CMS (16) and the COSMIN criteria. Previous studies have found that the QA-CMS has demonstrated excellent test-retest reliability (ICC >0.90), and this is consistent with the findings of this study. (17, 18, 44) The COSMIN was developed with an international committee and has gained acceptance for use in measurement reviews, although reliability has not been consistently good.(45) The QA-CMS was selected for its ease of use, track record of reliability and flexibility to deal with different types of measurement studies.

The strengths of the current study are the inclusion of articles from any language; to reduce the potential for bias since measures may perform differently across contexts. We found that there is insufficient volume of clinical measurement evidence to fully define all of the measurement properties of these tools, it was critical that we did not miss relevant studies. The weakness of our review limited the recommendations arising from this work and it should be considered when interpreting our findings. We could not calculate a pooled estimate for key psychometric properties because of the small number of studies and the heterogeneity in the studies. One of the co-authors of the paper is the developer of the PREE, and our critical appraisal tool which may have affected our reliability. However, this author was secondarily

involved and primary data extraction was performed by the first author and the last author. Since our findings were equally supportive of the PREE and pASES-e, there does not appear to be a bias in our findings or recommendations.

Future studies investigating the clinical measurement properties of the PREE and the pASES-e may consider the following recommendations in their design and reporting: 1) Use powerful statistical methods that are part of modern psychometric methods such as Rasch analysis to analyse the fit of the PROs to the model requirements, 2) provide clear justification of the sample size used 3) state clear hypotheses of what is expected for the properties being tested 4) calculate clinically important estimates like MCID, SEM etc. and 5) provide clear and specific conclusions that reflect the purpose of the review.

In conclusion, the PREE and the pASES-e have been shown to be valid, reliable and sensitive to change in both clinical and research settings based on high quality evidence. However, to improve the clinical utility, clinically important measurement estimates such as the SEM and MCID should be established. Future studies should focus on using modern psychometric methods to evaluate these two PROs against rigorous statistical models.

REFERENCES

1. Danis M, Solomon M. Providers, payers, the community, and patients are all obliged to get patient activation and engagement ethically right. Health Aff. 2013; 32: 401-7.

2. Small N, Bower P, Chew-Graham CA, Whalley D, Protheroe J. Patient empowerment in long-term conditions: Development and preliminary testing of a new measure. BMC Health Serv Res. 2013; 13: 263,6963-13-263.

3. Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: A review. J Clin Epidemiol. 1992; 45: 743-60.

4. Patient-rated elbow evaluation [homepage on the Internet]. Available from: http://www.srs-mcmaster.ca/Portals/20/pdf/research_resources/PREE.pdf.

5. MacDermid JC. Outcome evaluation in patients with elbow pathology: Issues in instrument development and evaluation. Journal of Hand Therapy. 2001; 14: 105-14.

6. King GJ, Richards RR, Zuckerman JD, Blasier R, Dillman C, Friedman RJ, et al. A standardized method for assessment of elbow function. research committee, american shoulder and elbow surgeons. J Shoulder Elbow Surg. 1999; 8: 351-4.

7. Liem IS, Kolling C, Marks M, Nelissen RG, Goldhahn J. Development of a score set to measure function and quality of life in patients suffering from elbow pathology. Arch Orthop Trauma Surg. 2012; 132: 831-7.

8. Hanyu T, Watanabe M, Masatomi T, Nishida K, Nakagawa T, Nishiura Y, et al. Reliability, validity, and responsiveness of the japanese version of the patient-rated elbow evaluation. Journal of Orthopaedic Science. 2013; 18: 712-9.

9. John M, Angst F, Awiszus F, King GJ, MacDermid JC, Simmen BR. The american shoulder and elbow surgeons elbow questionnaire: Cross-cultural adaptation into german and evaluation of its psychometric properties. J Hand Ther. 2010; 23: 301,13; quiz 314.

10. John M, Angst F, Pap G, Junge A, Mannion AF. Cross-cultural adaptation, reliability and validity of the patient rated elbow evaluation (PREE) for german-speaking patients. Clin Exp Rheumatol. 2007; 25: 195-205.

11. The B, Reininga IH, El Moumni M, Eygendaal D. Elbow-specific clinical rating systems: Extent of established validity, reliability, and responsiveness. J Shoulder Elbow Surg. 2013; 22: 1380-94.

12. Nourbakhsh E, Nugent R, Wang H, Cevik C, Nugent K. Medical literature searches: A comparison of PubMed and google scholar. Health Info Libr J. 2012; 29: 214-22.

13. Shariff SZ, Bejaimal SA, Sontrop JM, Iansavichus AV, Haynes RB, Weir MA, et al. Retrieving clinical evidence: A comparison of PubMed and google scholar for quick clinical searches. J Med Internet Res. 2013; 15: e164.

14. The patient rated elbow evaluation (PREE) - user manual [homepage on the Internet]. . June, 2010. Available from: http://www.srs-mcmaster.ca/Portals/20/pdf/research_resources/PREE%20User%20Manual_June%202010.pdf.

15. Liberati A, Altman DG, Tetzlaff J, Mulrow C, G\otzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. BMJ. 2009; 339.

16. MacDermid J. Quality appraisal for clinical measuremnet studies- evaluation form and guidelines. In: Law MC, MacDermid J, editors. Evidence-Based Rehabilitation :A Guide to Practice. 3rd ed. Thorofare, NJ: Slack; 2014. p. 325-38.

17. Roy JS, Desmeules F, MacDermid JC. Psychometric properties of presenteeism scales for musculoskeletal disorders: A systematic review. J Rehabil Med. 2011; 43: 23-31.

18. Roy JS, MacDermid JC, Woodhouse LJ. Measuring shoulder function: A systematic review of four questionnaires. Arthritis Rheum. 2009; 61: 623-32.

19. Sim J, Wright CC. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. Phys Ther. 2005; 85: 257-68.

20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33: 159-74.

21. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: What are the necessary measurement properties?. J Clin Epidemiol. 1992; 45: 1341-5.

22. Grayson K, Rust R. Interrater reliability. Journal of Consumer Psychology. 2001; 10: 71-3.

23. Wong SP, McGraw KO. Confidence intervals and F tests for intraclass correlations based on three-way random effects models. Educational and Psychological Measurement. 1999; 59: 270-88.

24. Wright BD. Interpreting reliabilities. Rasch Measurement Transactions. 1998; 11: 602.

25. Vincent JI, King GJW, Grewal R. Validity and sensitivity to change of patient-reported pain and disability measures for elbow pathologies. J Orthop Sports Phys Ther. 2013; 43: 263-74.

26. Schmidt T, Stangl R. Indications, complications and results of elbow arthroscopy: A retrospective assessment with correlation analysis of the german version of the patient rated elbow evaluation (PREE) score. Obere Extremitat. 2013; 8: 220.

27. Angst F, Goldhahn J, Drerup S, Kolling C, Aeschlimann A, Simmen BR, et al. Responsiveness of five outcome measurement instruments in total elbow arthroplasty. Arthritis Care and Research. 2012; 64: 1749-55.

28. Angst F, John M, Pap G, Mannion AF, Herren DB, Flury M, et al. Comprehensive assessment of clinical outcome and quality of life after total elbow arthroplasty. Arthritis Care and Research. 2005; 53: 73-82.

29. Vincent JI, MacDermid JC, King GJ. Response to: Critical assessment of patient-reported outcome measures. J Orthop Sports Phys Ther. 2013; 43: 513-4.

30. Towards a common language for functioning, disability and health ICF [homepage on the Internet. 2002. Available from: http://www.who.int/classifications/icf/training/icfbeginnersguide.pdf.

31. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: Are available health status surveys adequate?. Qual Life Res. 1995; 4: 293-307.

32. Chung KC, Hamill JB, Kim HM, Walters MR, Wilkins EG. Predictors of patient satisfaction in an outpatient plastic surgery clinic. Ann Plast Surg. 1999; 42: 56-60.

33. Roy JS, Macdermid JC, Goel D, Faber KJ, Athwal GS, Drosdowech DS. What is a successful outcome following reverse total shoulder arthroplasty?. Open Orthop J. 2010; 4: 157-63.

34. Linacre JM. Detecting multidimensionality: Which residual data-type works best?. J Outcome Meas. 1998; 2: 266-83.

35. Smith EV,Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas. 2002; 3: 205-31.

36. Andrich D. Rasch models for measurement. Newbury Park: Sage Publications; 1988.

37. DeMars C. Item response theory. Oxford ; New York: Oxford University Press; 2010.

38. Brown TA. Confirmatory factor analysis for applied research. Guilford Publications, Incorporated; 2006.

39. Bland JM, Altman DG. Measuring agreement in method comparison studies. Statistical Methods in Medical Research. 1999; 8: 135-60.

40. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. statistics and strategies for evaluation. Control Clin Trials. 1991; 12: 142S-58S.

41. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care. 1989; 27: S178-89.

42. Guyatt G, Walter S, Norman G. Measuring change over time: Assessing the usefulness of evaluative instruments. J Chronic Dis. 1987; 40: 171-8.

43. Sanderson S, Tatt ID, Higgins JPT. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography. Int J Epidemiol. 2007; 36: 666-76.

44. MacDermid JC, Walton DM, Avery S, Blanchard A, Etruw E, McAlpine C, et al. Measurement properties of the neck disability index: A systematic review. J Orthop Sports Phys Ther. 2009; 39: 400-17.

45. Mokkink LB, Terwee CB, Gibbons E, Stratford PW, Alonso J, Patrick DL, et al. Inter-rater agreement and reliability of the COSMIN (COnsensus-based standards for the selection of health status measurement instruments) checklist. BMC Med Res Methodol. 2010; 10: 82,2288-10-82.

Table 1: Summary table describing the studies included in the systematic review

Study	quality score	PRO	*Population*	n	Clinical measurement properties measured
Angst et al (2012)(32)	95.83	ASES-e; PREE	*Total elbow arthroplasty (RA and post-trauma cases); M:F - 19:46; Age [mean(SD)] = 61.9(13.0)*	65	Responsiveness (SRM and ES); ROC curve analysis; test-retest reliability; Sensitivity and specificity
Vincent et al (2013)(30)	87.5	PREE; ASES-e	*Various elbow surgeries M:F – 84:44; Age [mean(SD)] = 46.5 (12.8)*	128	Internal consistency; construct validity; longitudinal validity; responsiveness (SRM and ES), factor analysis
Hanyu et al (2013)(13)	83.33	PREE	*Various elbow pathologies M:F – 39:35; Age [mean(SD)] = 46.7 (20.7)*	74	Translation and cross-cultural adaptation; test-retest reliability; internal consistency; construct validity; responsiveness, factor analysis
MacDermid et al (2001)(10)	79.17	PREE; ASES-e	*Elbow trauma (Operative and non-operative cases); Infection; OA; RA; lateral epicondylitis; M:F - 33:37; Age [mean(SD)] = 49(16)*	70	Test-retest reliability Construct validity Divergent validity
John et al (2010)(14)	79.17	ASES-e	*Total elbow arthroplasty (RA and post-trauma cases); M:F - 22:53; Age [mean(SD)] = 64.1(13.3)*	75	Translation and cross-cultural adaptation; test-retest reliability; internal consistency; construct validity; Divergent validity; Floor and ceiling
Angst et al (2005)(33)	75	PREE; ASES-e	*Total elbow arthroplasty (RA and post-trauma cases; M:F - 23:56; Age[mean(SD)] = 64.1(13.3)*	79	Construct validity; exploratory factor analysis; Known group validity
John et al (2007)(15)	75	PREE	*Elbow prosthesis; M:F - 19:37; Age [mean(SD)] = 63.7(11.4)*	56	Translation and cross-cultural adaptation; test-retest reliability; internal consistency; construct validity; Divergent validity
Schimdt and Stangl (2013)(31)	70.83	PREE	*Various elbow pathologies M:F – 39:16; Age mean(Range)] = 46 (13 – 71)*	55	Construct validity
Liem et al (2012)(12)	54.16	PREE	*Various elbow pathologies M:F 33:33; Age [mean(Range)] = 49.99 (19.4 – 72.5)*	66	Construct validity

PREE- Patient rated elbow evaluation; ASES-e – American shoulder and elbow surgeons – elbow questionnaire; M- Male; F – Female; SD – Standard deviation

Table 2: Quality of studies included in the systematic review (arranged according to study quality)

Study	Item evaluation score for each criteria on the MacDermid quality assessment tool (Min=0; Max=2)												Total score	Quality score (%)
	1	2	3	4	5	6	7	8	9	10	11	12		
Angst et al (2012)(32)	2	2	2	2	2	1	2	2	2	2	2	2	23	95.83
Vincent et al (2013)(30)	2	2	2	2	1	2	2	1	2	2	2	1	21	87.50
Hanyu et al (2013)(13)	2	2	2	2	0	1	2	2	2	2	2	1	20	83.33
MacDermid et al (2001)(10)	2	1	2	2	0	1	2	2	2	2	2	1	19	79.17
John et al (2010)(14)	2	1	2	2	0	1	2	2	2	2	2	1	19	79.17
Angst et al (2005)(33)	2	1	2	2	0	N/A	2	2	2	2	1	2	18	75.00
John et al (2007)(15)	2	1	1	2	0	1	2	2	2	2	2	1	18	75.00
Schimdt & Stangl (2013)(31)	1	2	2	1	0	N/A	2	2	2	2	2	1	17	70.83
Liem et al (2012)(12)	1	2	2	0	0	N/A	1	1	2	2	1	1	13	54.16

N/A – Not applicable to the study

Disagreement in 30 instances 78/108*100= 72.22 before consensus

Table 3: Summary of validity properties, cross cultural adaptations and administrative and responder burden of the PREE and the pASES-e

Clinical measurement property	Data extracted	
	PREE	**pASES-e**
Content Validity (includes analysis of questions, Floor and Ceiling effects, missing items)	-All questions were easily understood and completed by all patients.(33) -No patients had difficulty completing the questionnaire. (Japanese version)(13) **<u>Floor and Ceiling effects</u>** **Pain subscale**: Floor – 0%; Ceiling – 15%(33) Floor – 1%; Ceiling – 3% (Japanese version)(13); Floor – 0%; Ceiling – 14.3% (German version)(15) **Function subscale:** Floor – 0%; Ceiling – 4%(33) Floor – 1%; Ceiling – 4% (Japanese version)(13) Floor – 0%; Ceiling – 5.3% (German version)(15) **Total score:** Floor – 0%; Ceiling – 4%(33) Floor – 1%; Ceiling – 0% (Japanese version)(13) Floor – 0%; Ceiling – 5.3% (German version) (15)	-All questions were easily understood and completed by all patients.(33) **<u>Floor and Ceiling effects</u>** **Pain subscale**: Floor – 0%; Ceiling – 14%(33) **Function affected subscale:** Floor – 0%; Ceiling – 4%(33) **Satisfaction subscale:** Floor – 3%; Ceiling – 43%(33) **Total score:** Floor – 0%; Ceiling – 3%(33)
Factor structure	-Principal component analysis with varimax rotation supported a 4 factor model with only one cross loading.(30) -PREE pain has a unidimensional structure; PREE function has a bi-dimensional structure.(Japanese version)(13)	Principal component analysis with varimax rotation supported a 4 factor model with 2 cross loadings.(30)
Construct	**<u>Reported high Correlations (> 0.70):</u>**	**<u>Reported high Correlations (></u>**

convergent validity	**Pain subscale:** PREE Function score(12, 13) PREE total score(12, 13, 30) pASES-e pain subscale(10, 30) Quick DASH(31) DASH total (Japanese version)(13) **Function subscale:** PREE pain score (12, 13) PREE total score(12, 30) DASH total score(13, 15, 30) Quick DASH(12, 31) DASH(10) **Total score:** PREE pain (12, 13, 30) PREE function(12, 13, 30) pASES-e pain(30) DASH total (10, 13, 15, 30) Quick DASH (12, 31) pASES-e total score(33) **Reported moderate Correlations (0.30 to 0.70):** **Pain subscale:** PREE function (30) pASES-e function affected (30) DASH total(15, 30) SF-36 PCS score(10, 15) Quick DASH (12) **Function subscale:** PREE pain (30) pASES-e pain(30) pASES-e function affected (30) SF-36 PCS score (10, 13, 15, 30, 33) cmASES(15) **Total score:** pASES-e function affected (30) SF-36 PCS score (10, 13, 15, 30) cmASES (33)	**0.70):** **Pain subscale:** PREE pain subscale (10, 30) PREE total score(14, 30) DASH (10) **Function affected subscale:** DASH (14) SF-36 Physical component summary (14) PREE function score (14) PREE total score(14) **Satisfaction subscale:** None **Total score:** DASH(14) PREE function score(14) PREE total score(14) **Reported moderate Correlations (0.30 to 0.70):** **Pain subscale:** PREE function subscale (30) pASES-e function affected (30) DASH total score (14, 30) SF-36 PCS score(10, 14) cmASES (15) **Function affected subscale:** PREE pain (30) PREE function (30) PREE total score (30) pASES-e pain(30) DASH total score (30) SF-36 PCS score (10, 30) cmASES(15) **Satisfaction subscale:** SF-36 PCS score(10, 15) DASH total score(15) **Total score:** cmASES (15, 33) SF-36 PCS score(10, 15, 33)
Construct divergent validity	**Pain subscale:** SF-36 Mental component summary(13) **Function subscale with:** SF-36 Mental component summary(13,	**Pain subscale:** SF-36 Mental component summary(15) **Function affected subscale with:** SF-36 Mental component summary(15,

	30, 33) **Total score with:** SF-36 Mental component summary(10, 12, 13, 15, 30)	30) **Satisfaction subscale with:** SF-36 Mental component summary(15) **Total score with:** SF-36 Mental component summary(15, 33)
Cross-cultural adaptation	• German version.(15) Translated with inputs from the developer and found psychometric properties that was equivalent to the original English version. • Japanese version.(13) Found the adapted version to exhibit clinical measurement properties equivalent to the original English version.	• German version.(14) Found clinical measurement properties that were equivalent to the original English version.
Response burden	• Time to complete: 3 minutes (33) • Clarity of items: Was easily understood.(32) • Reading level: Flesch level of 6.3 which is equivalent to Grade 6 reading level. (34)	• Time to complete: 3 minutes(33) • Clarity of items: Was easily understood.(33)(32)

PREE – Patient Rated Elbow Evaluation; pASES-e – self-report section of the American Shoulder and Elbow Surgeons – elbow form; SF-36

Table 4: Summary of reliability indices for the PREE and the pASES-e

Clinical measurement property	Data extracted	
	PREE	**pASES-e**
Test-retest reliability (ICC)	**Pain subscale:** 0.92 (Japanese version)(13) 0.90 in RA cases (Japanese version) (13) 0.88 (95% CI 0.78 – 0.94) (10) 0.73 (German Version)(15) **Function Subscale:** 0.93 (Japanese version) (13) 0.86 in RA cases (Japanese version) (13) 0.89 (95% CI 0.79 – 0.94) (10) 0.82 (German Version) (15) **Total score:** 0.94 (Japanese version) (13) 0.90 in RA cases (Japanese version) (13) 0.95 (95% CI 0.86 – 0.98) (10) 0.80 (German Version) (15)	**Pain subscale:** 0.90 (95% CI 0.82 – 0.94) (German Version)(14) 0.89 (95% CI 0.80 – 0.94) (10) **Function Subscale(Affected):** 0.87 (95% CI 0.78 – 0.92) (German Version) (14)0.79 (95% CI 0.66 – 0.88) (10) **Function Subscale(Unaffected):** 0.74 (95% CI 0.53 – 0.87) (German Version) (14) 0.64 (95% CI 0.42 – 0.78) (10) **Satisfaction subscale:** 0.84 (95% CI 0.71 – 0.91) (10) **Total score (affected):** 0.93 (95% CI 0.88 – 0.96) (German Version) (14) **Total score (unaffected):** 0.92 (95% CI 0.84 – 0.96) (German Version) (14)
Standard Error of Measurement (SEM)	None	**Pain subscale:** 8.94 (95% C.I -21.23 to 14.74) (German Version(14) **Function Subscale (Affected):** 9.04 (95% C.I -20.83 to 15.40) (German Version) (14)**Function Subscale(Unaffected):** 11.62 (95% CI -23.74 to 23.28) (German Version) (14) **Satisfaction subscale:** 1.32 (95% C.I -2.05 to 3.25) (German Version) (14) **Total score (Affected):** 6.23 (95% C.I -15.8 to 9.88) (German Version) (14) **Total score (Unaffected):** 6.61 (95% C.I -14.70 to 12.02) (German Version)(14)

MDC $_{95\%}$	None	**Pain subscale:** 23.51(German Version) (14) **Function Subscale (Affected):** 23.42 (German Version)(14) **Function Subscale (Unaffected):** 31.43 (German Version) (14) **Satisfaction subscale:** 3.85 (German Version) (14) **Total score (Affected):** 16.05 (German Version) (14) **Total score (Unaffected):** 18.27 (German Version)(14)
Mean retest difference (Bland and Altman plot)	None	**Total Score:** Affected side mean difference and narrowest 95% CI -2.6 $\pm$ 16.05 (German Version) (14)
Internal consistency (Cronbach's alpha)	**Pain subscale:** 0.92(95% CI 0.88 – 0.93) (Japanese version) (13) 0.93 (German Version) (15) **Function Subscale:** 0.97 (95% CI 0.97 – 0.97) (Japanese version) (13) 0.95 (German Version) (15) **Total score:** 0.95 (95% CI 0.94 – 0.96)(30) 0.97 (95% CI 0.97 – 0.97) (Japanese version) (13) 0.96 (German Version) (15)	**Pain subscale:** 0.91(German Version) (14) **Function Subscale (affected):** 0.93 (German Version) (14) **Function Subscale (unaffected):** 0.94 (German Version) (14) **Total score (affected):** 0.93 (95% CI 0.91 – 0.94) (30) 0.90 (German Version)(14) **Total score (unaffected):** 0.90 (German Version) (14)

PREE – Patient Rated Elbow Evaluation; pASES-e – self-report section of the American Shoulder and Elbow Surgeons – elbow form; ICC – Intra-class correlation co-efficient; RA – Rheumatoid arthritis

Table 5: Summary of responsiveness properties of the PREE and the pASES-e

Clinical measurement property	Data extracted	
	PREE	**pASES-e**
Effect size (ES)	**Pain sub scale:** 1.7 (32) 1.4 (95%CI 1.2 – 1.6) (30) 1.32 (Japanese version) (13) **Function sub scale:** 0.99 (32) 1.4 (95%CI 1.3 – 1.6) (30) 0.86 (Japanese version) (13) **PREE total:** 1.5 (32) 1.7 (95%CI 1.5 – 1.8) (30) 1.12 (Japanese version) (13)	**Pain sub scale:** 1.55 (32) 1.4 (95%CI 1.2 – 1.6) (30) **Function sub scale:** 0.77 (32) 1.3 (95%CI 1.1 – 1.5) (30) **Total Score:** 1.32 (32)
Standardized Response Mean (SRM)	**Pain sub scale:** 1.27 (32) 1.3 (95%CI 1.0 – 1.4) (30) 1.31 (Japanese version) (13) **Function sub scale:** 0.97 (32) 1.5 (95%CI 1.4 – 1.7) (30) 1.02 (Japanese version) (13) **PREE total**: 1.37 (32) 1.6 (95%CI 1.4 – 1.8) (30) 1.28 (Japanese version) (13)	**Pain sub scale:** 1.15 (32) 1.2 (95%CI 1.0 – 1.4) (30) **Function sub scale:** 0.75 (32) (95%CI 0.9 -1.3) (30) **Total Score:** 1.17 (32)
Longitudinal Validity correlation with other change scores	**Pain sub scale change score:** With pASES-e pain: 0.85 ($p<0.01$) (30) With pASES-e Function: 0.24 ($p<0.01$) (30) With DASH:0.41($p<0.01$) (30) **Function sub scale change score:** With pASES-e pain: 0.41 ($p<0.01$) (30) With pASES-e Function: 0.40 ($p<0.01$) (30)	**Pain sub scale change score:** With PREE pain: 0.85 ($p<0.01$) (30) With PREE Function: 0.41 ($p<0.01$) (30) With PREE total score: 0.74 ($p<0.01$) (30) With DASH:0.46($p<0.01$)(30) **Function sub scale change score:** With PREE pain: 0.24 ($p<0.01$) (30) With PREE Function: 0.40 ($p<0.01$) (30)

	With DASH:0.62 ($p<0.01$) (30) **PREE Total change score:** With pASES-e pain: 0.74 ($p<0.01$) (30) With pASES-e Function: 0.33 ($p<0.01$)(30) With DASH:0.62 ($p<0.01$) (30)	With PREE total score: 0.33 ($p<0.01$) (30) With DASH:0.23($p<0.01$) (30)
Area Under the Curve (AUC)	**Pain sub scale:** 0.65 (95% CI 0.51 – 0.78); Sensitivity = 0.67; Specificity = 0.71 (32) **Function sub scale:** 0.62 (95% CI 0.48 – 0.75); Sensitivity = 0.83*; Specificity = 0.47* (32) **Total Score:** 0.68 (95% CI 0.57 – 0.81); Sensitivity = 0.63; Specificity = 0.71 (32)	**Pain sub scale:** 0.64 (95% CI 0.51 – 0.78); Sensitivity = 0.64; Specificity = 0.71 (32) **Function sub scale:** 0.62 (95% CI 0.47 – 0.78); Sensitivity = 0.58*; Specificity = 0.63* (32) **Total Score:** 0.67 (95% CI 0.53 – 0.80); Sensitivity = 0.65; Specificity = 0.69 (32)

* - not significant; PREE – Patient Rated Elbow Evaluation; pASES-e – self-report section of the American Shoulder and Elbow Surgeons – elbow form; DASH – Disabilities of the Arm Shoulder and the Hand questionnaire

Figure 1: Systematic review evidence flowchart based on PRISMA guidelines

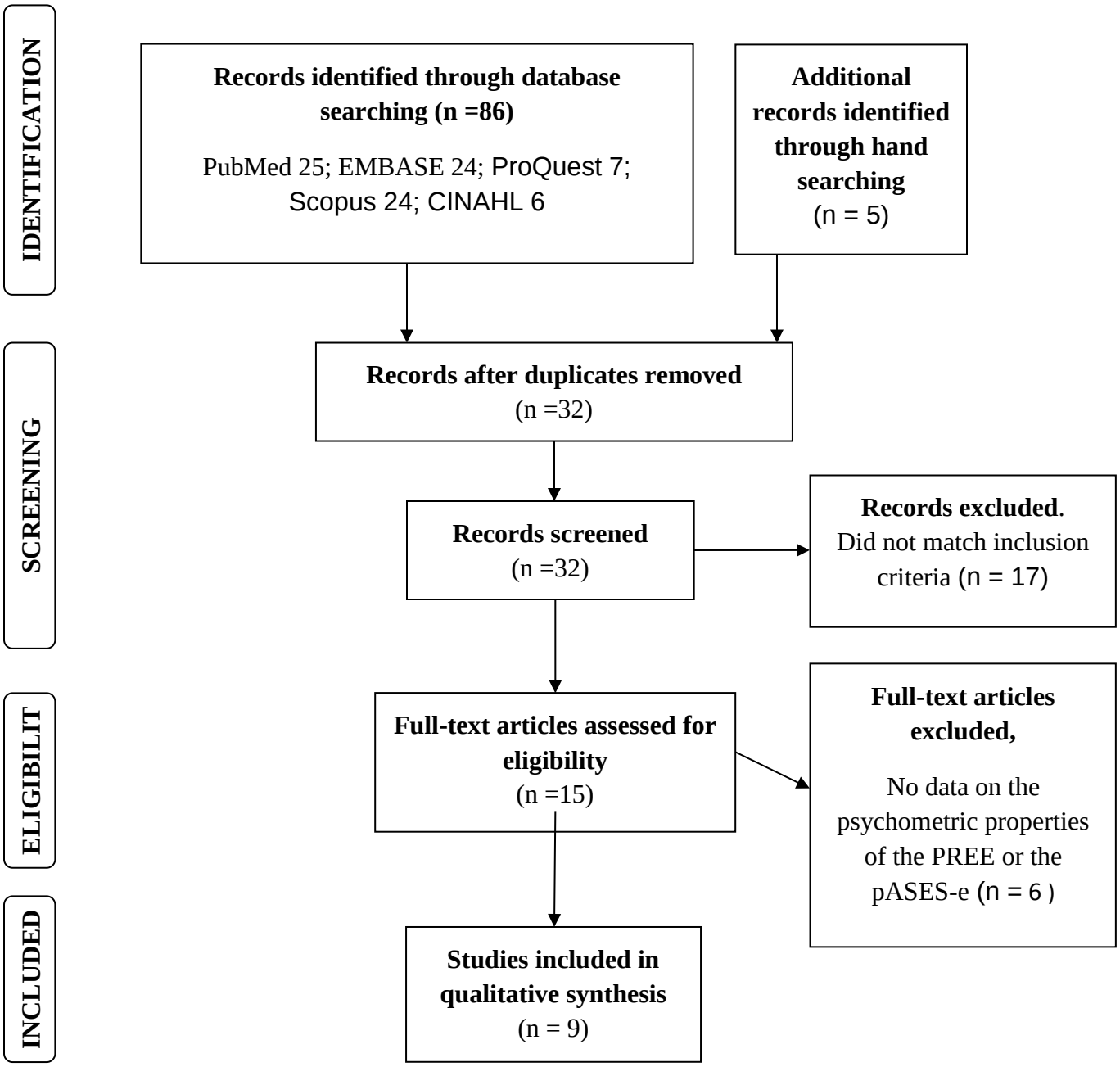


Figure legend: CINAHL - Cumulative Index to Nursing and Allied Health Literature; PRISMA - Preferred Reporting Items for Systematic Reviews and Meta-Analyses

CHAPTER 4: LINKING OF THE PATIENT RATED ELBOW EVALUATION (PREE) AND THE AMERICAN SHOULDER AND ELBOW SURGEONS – ELBOW QUESTIONNAIRE (pASES-e) TO THE INTERNATIONAL CLASSIFICATION OF FUNCTIONING DISABILITY AND HEALTH (ICF) AND HAND CORE SETS [1]

ABSTRACT

Study Design: Content analysis-ICF linking

Introduction: The Patient Rated elbow evaluation (PREE) and the self-report section of the American Shoulder and Elbow Surgeons society – Elbow form (ASES-e) are the two commonly used elbow pain and disability self-report measures (PROs). The content of these questions have never been analysed in light of the International Classification of Functioning Disability and Health (ICF) which is the current standard to describe health and health related states.

Purpose: The purposes of this study were to analyse the conceptual basis of the PREE and the ASES-e by linking the meaningful concepts in these PROs to the ICF using standardized linking rules and to determine the extent to which the ICF core set for hand conditions cover the content of elbow questionnaires using summary ICF linkage indicators.

Methods: Two raters linked the two PROs to the ICF using the linking rules proposed by Cieza and colleagues. Percentage agreement was calculated between the raters. Summary linkage indicators proposed by MacDermid were used to estimate the extent to which the ICF core set for hand conditions cover the content of the elbow questionnaires.

Results: All the items of the PREE (*Measure to ICF linkage-100%)* and all but one item of the pASES-e (*Measure to ICF linkage-95%)* were linked to the ICF. The satisfaction item on the ASES-e was not-covered by the ICF. Percentage agreement on linking between the raters was 96% and 95% for the PREE and the pASES-e respectively. The unique linkage of the PREE and the pASES-e to the unique codes on the brief and comprehensive core set lower than absolute

---

A version of this work has been submitted for journal publication**: Vincent, JI,** MacDermid, JC, King GJW, Grewal, R. Linking of the Patient Rated Elbow Evaluation (PREE) and the American Shoulder and Elbow Surgeons – Elbow questionnaire (pASES-e) to the International Classification of Functioning Disability and Health (ICF) and Hand Core Sets. Journal of Hand Therapy (2014)

linkage to the core set for hand conditions. The PROs represented less than 20% of the comprehensive core set and more than 70% of the brief core set. While for the unique core set disability representation the *2* measures represented 100% brief core set unique disability codes and less than 35% of the comprehensive core set unique disability codes.

Conclusion: The PREE and the ASES-e are aligned with the ICF framework and the core sets for hand conditions. The ICF Core Set devised for hand conditions may also be useful for elbow conditions..

Level of evidence: 1b

Keywords: ICF, PREE, pASES-e, Content analysis, ICF core set for hand conditions, ICF linking, Linkage indicators

INTRODUCTION

Patient centered outcome research and patient reported outcomes (PRO) have become an integral component in the field of health care. These measures should undergo rigorous validation before they are accepted into clinical practice. The Patient Rated elbow evaluation (PREE)[1] and the self-report section of the American Shoulder and Elbow Surgeons society – Elbow form (pASES-e)[2] are currently among the most commonly used elbow self-report pain and disability measures used in clinical practice and research. While multiple studies[1,3-5] support the psychometric properties of these tools there has been less attention to content validity.

Content analysis is a research technique for making replicable and valid inferences from texts to the contexts of their use.[6] Content validity of questionnaires is sometimes assessed by expert review, but is often not structured and sparsely reported. Formal analysis of the content of items of PRO should be a fundamental step in defining the conceptual domain of measures. However content validity is often sparsely addressed in the literature, with informal review by experts being common.[7,8]

The International Classification of Functioning Disability and Health (ICF) is a framework for classifying health and health-related states.[9] The ICF enables the transfer and interpretation of data across borders and disciplines by providing a common language and a hierarchical coding system to describe health and disability. Since PROs are often designed to measure disability, coding the content of items using the ICF can provide a structured evaluation of content. It is recommended that meaningful concepts in the questionnaires should be linked to the ICF codes using established linking guidelines to provide a structured analysis of the content of PRO.[10-12]

The PREE and the ASES-e were developed using a process where items were generated from pre-existing questionnaires and/or patient interviews. As such, the ICF conceptual framework was not taken into consideration while developing these PROs. Both these PROs have been recommended as key measures for assessing outcomes in elbow disorders.[13,14] Cross-cultural validation often evaluates whether questionnaire items can be developed to provide equivalent estimates of disability in a different language or cultural context. For example, the PREE has been translated into German and Japanese, while the pASES-e has been translated into

German.[15-17] Linking PROs to the ICF can assist with cross-cultural validation or assessment of equivalence across translated measures.

The core set for hand conditions[18] was developed by the ICF Research Branch in collaboration with other organisations to help clinicians who are specialised in conditions of the 'hand'. In this context, 'hand' conditions refer to problems located directly at the hand such as, carpal tunnel syndrome, osteoarthritis of the hand or finger joints, amputation of fingers, Dupuytren's, hand injuries etc. As well, 'hand' conditions could also refer to conditions which originate at a different part of the body but affect the hand such as stroke, Parkinson's disease and multiple sclerosis.[18] Functionally, the shoulder provides a stable base and the elbow allows placement of the hand in the space for it to function. Hence, clinicians often consider the function of the upper extremity as a unit. Currently there is no specific ICF core set to help with assessment of upper extremity conditions, although it might be considered that those established for the hand would be relevant to the entire upper extremity.

PURPOSE

The purpose of this study was to analyse the conceptual basis of the PREE and the ASES-e by linking the meaningful concepts in these PRO's to the ICF following standardized linking rules and also to assess the extent to which the ICF core set for hand conditions cover the content of elbow questionnaires by obtaining the ICF linkage indicators.

METHODS

Description of measures:

Patient rated elbow evaluation (PREE): The PREE is a 20 item joint specific pain and disability PRO of the elbow.[1] The 20 items are grouped into 3 sections. The first section contains 5 items on pain. The second and third sections are on specific activities (11 items) and usual activities (4 items) respectively. Each of the items are scored on an 11 point scale ranging from 0 – 10. The total score is out of 100 with higher scores indicating greater disability. The pain and function items are weighted equally in the total score.[19] The clinical measurement properties of the PREE have been previously examined. The PREE has been shown to exhibit moderate to high correlations with the ASES-e, the DASH, MEPI and the physical component summary

score of the SF-36.[1,3,5] The PREE has also been shown to exhibit high test-retest reliability(>0.80)[1,4] and excellent internal consistency (>0.90).[5] It has also demonstrated high sensitivity to change with large standardized response means (SRM) and effect sizes (ES) (>1).[4,5] It has also been translated into German[16] and Japanese.[17]

American Shoulder and Elbow Surgeons society – Elbow form (pASES-e): The pASES-e is an 18 item self-report questionnaire designed to measure pain and disability arising from elbow disorders.[2] It has 3 sections on pain, function and satisfaction. The pain section has five items and is scored on a 0 – 10 scale. The function section has 12 items scored on a 0 – 3 scale. Patients are asked to rate both their affected and the unaffected side. The satisfaction section has one item scored on a 0 – 10 scale. The clinical measurement properties of the pASES-e are also well established. It has been shown that the pASES-e exhibits moderate correlations with the PREE and the DASH.[1,5] It has also been shown to have high internal consistency (>0.80)[5] and test-retest reliability(>0.90)[1,4] and is highly sensitive to change with a large SRM and ES (>1).[4,5] It has been translated and cross-culturally adapted into German.[15]

Content analysis of the PREE and the pASES-e using the ICF

ICF coding*:* The linking rules proposed by Cieza and colleagues were used to link the meaningful concepts of the PREE and the pASES-e to the ICF.[10-12] According to these rules, each item of a health-status measure should be linked to the most precise ICF category; if a single item encompasses different constructs, the information in each construct should be linked; if the content of an item is more general than the corresponding ICF category, then the code of the higher level is linked; if the information provided by an item is not sufficient for making a decision about which ICF category the item should be linked to, this item is assigned nd (not definable); if an item is not contained in the ICF classification, then this item is assigned nc (not covered by ICF).[10-12] In the current study, [11] two independent raters identified meaningful concepts in each of the items on the PREE and the pASES-e and then linked these concepts using the ICF codes independently. Consensus was measured between the raters by percentage agreement (A) by dividing the observed agreement (O) by the possible agreement (P). [20]

$$\text{Percentage agreement (A)} = \frac{\text{Observed agreement (O)}}{\text{Possible agreement (P)}} \times 100\%$$

***Measure to ICF linkage*:** To measure the breadth of association between the concepts in the measures and the ICF, a 'measure to ICF linkage'[21] criteria was used. The 'measure to ICF linkage' was defined as the percentage of items from a measure that can be linked to ICF codes. This percentage represents the extent to which the content in a measure could be expressed using the ICF codes and was calculated using the following formula:

$$\text{Measure to ICF linkage} = \frac{\text{The number of items linked to at least 1 ICF code}}{\text{Total number of items on the measure}} \times 100\%$$

Assessing the extent to which the ICF core set for hand conditions covers the content of the PREE and the pASES-e

ICF core set linkage indicators: To define the depth and breadth of linkage between items of the PREE and pASES-e and the ICF core sets for hand conditions, we used linkage indicators developed by the second author (JMD).[22] These indicators are simple equations that will enable us to understand how well the content of the items of the PREE and pASES-e linked to the codes in the ICF core sets for hand conditions. The ICF core set for hand conditions has two parts; a comprehensive core set and a brief core set. The comprehensive core set includes 117 ICF categories which includes 38 disability codes that can be taken into account when conducting a comprehensive, multidisciplinary assessment. The brief core set consists of 23 categories which includes 8 disability codes from the comprehensive core set to enable quick assessment of individual patients in routine clinical practice. The following equations were developed as indicators:

1) **Measure to (brief or comprehensive) core set absolute linkage** was defined as the percentage of items from a measure that could be linked to ICF codes on the core set (brief or comprehensive) and is calculated using the following equation:

$$\text{Measure to (Brief or Comprehensive) Core Set Absolute Linkage} = \frac{\text{Number of items linked to a code(s) appearing in the CoreSet}}{\text{Total number of items on the measure}} \times 100\%$$

2) Measure to (brief or comprehensive) core set unique linkage was defined as the percentage of measure's items that could be linked to unique ICF codes and represented the extent to which the items of a measure represent different content indicated by the core set. Once an item was

coded to a core set item, additional items that coded to that same code were not counted again. This indicator was calculated using the following equation:

$$\text{Measure to (Brief or Comprehensive) Core Set Unique Linkage} = \frac{\text{Number of items that are linked to Unique codes in Core Set}}{\text{Total number of items on the scale}} \times 100$$

3) Core set representation was defined as the percentage of unique core set codes that appeared when the measure's items were linked to ICF codes. This represents the extent to which the entire scope of the content defined by the core set was represented on the measure. This indicator was calculated using the following equation:

$$\text{Core Set Representation} = \frac{\text{Number of unique ICF codes from the measure that appear in the CoreSet}}{\text{Total number of codes on the (Brief or Comprehensive)CoreSet}} \times 100\%$$

4) Core Set Unique Disability Representation was defined as the percentage of unique core set disability codes that appeared when the measure's items were linked to ICF codes. For PROs that were designed to measure disability, it is be important to determine the extent to which they measure this aspect of content. This represented the extent to which the disability codes defined by the core set were represented on the measure. Once an item was coded to a core set disability code, additional items that coded to that same code were not counted again. This indicator was calculated using the following equation:

$$\text{Core Set Unique Disability Representation} = \frac{\text{Number of unique (d)codes from the measure that appear in the Core Set}}{\text{Total number of disability codes on the (Brief or Comprehensive)CoreSet}} \times 100\%$$

RESULTS

Analysing the conceptual basis of the PREE and the pASES-e in light of the ICF

Measure to ICF linkage:

The PREE and the pASES-e covered the important ICF domains of body function / structure; activity limitation and participation restriction while environmental factors domain was not covered. Twenty one categories (Body structure/function – 5; Activity limitation/participation restriction – 16) were required to link the items of the PREE to the ICF giving it a percentage score of 100%. (See Figure 1) The pASES-e had a slightly lower score (94%) than the PREE. (See Figure 1) The last item of the pASES-e "Are you satisfied with your elbow surgery?" was not-codable. On the whole, 15 categories (Body structures/functions – 5; Activity limitation and participation limitation – 10) were used to link the items of the pASES-e to the ICF.

Agreement between raters:

The raters agreed with each other on most occasions, resulting in a percentage agreement of 96% for the PREE and 95% for the pASES-e. (See Figure 1)

Assessing the extent to which the ICF core set for hand conditions cover the content of the PREE and the pASES-e

1. Measure to (Brief or Comprehensive) Core Set Absolute Linkage

All the items of the PREE and all but one item of the pASES-e were linked to the categories in the comprehensive core set for hand conditions resulting in an absolute linkage score of 100% and 95% respectively. (See Figure 2) The absolute linkage score to the brief core set was 85% for the PREE and 84% for the pASES-e (See Figure 2). The categories that were used to code 3 items of the PREE from The Specific Function Subscale, namely, "Use my arm to rise from the chair" and "Use a telephone" and "Recreational activities"; and 2 items from the pASES-e "Rise from chair pushing with arm" and "Do usual sport", were not part of The Brief Core Set for Hand Conditions.

2. Measure to (Brief or Comprehensive) Core Set Unique Linkage

The unique linkage of the PREE and the pASES-e to the unique codes on the comprehensive core set was 65% and 68% respectively. (See Figure 2) A similar trend was observed for the unique linkage of the 2 PROs to the brief core set for hand conditions. (PREE – 50%; pASES-e – 58%) (See Figure 2)

3. Core Set Representation

Both the PREE and the pASES-e represented less than 20% of the content of the comprehensive core set while representing about 74% and 70% of the content of the brief core set. (See Figure 2)

4. Core Set Unique Disability Representation

The PREE and the pASES-e achieved a disability representation score of 32% and 26% respectively when linked to the disability content of the comprehensive core set (See Figure 2). When the unique core set disability content representation was analysed, both the PREE and the pASES-e achieved a score of 100% (See Figure 2).

DISCUSSION

Although the PREE and the pASES-e were developed without explicit use of an ICF framework, this study has established that the concepts of impairment and disability appearing on these two scales are well aligned with ICF concepts. These two measures have also exhibited strong linkage to a subset of ICF codes contained on the Hand Core Sets. This supports the content validity of both the PREE and the pASES-e, since to measure symptoms (impairments) and function in people with elbow conditions. The linkage indicators developed by MacDermid[22] allowed us to summarize how the measures reflected content contained in the core set for hand conditions and suggested that the core set is appropriate for guiding assessment of disability in people with disorders of the elbow. Since the upper extremity often functions as a unit in accomplishing life tasks, this suitability is understandable. These summary measures may be helpful in comparing measures in future studies as it can be challenging to compare codes without summary indicators.

The content analysis performed in this current study provided a mechanism to compare the two measures. This is used by clinicians to measure based on their content and how it matches with their specific clinical need. Both the measures needed the same codes except for 3 instances. (See table 3, 4)The fact that both the PREE and the pASES-e were linked using almost the same categories under the important ICF domains of body function/structure, activity limitation and participation restriction and indicated that both the PROs are aligned to the ICF and are equally good at measuring what they are intended to measure. Clinicians can use these measures interchangeably depending on the level of comfort. Since the content is similar it is not surprising that they had correlated well ($r = 0.92$; $p<0.001$) in a previous study which measured the clinical measurement properties of these two measures.[3,5]

The pASES-e has three separate subscales and does not provide a total score. Previous studies have questioned the inclusion of the ASES-e item on satisfaction from surgery as an outcome indicator, since satisfaction can be affected by both process and outcome.[1,5] In the current study, the item on satisfaction from surgery was not codable using the ICF providing further support that its content does not align with the remaining items -or the concept of disability. This does however support the decision to not create a total score from the three subscales of the pASESe, since inclusion of satisfaction might be expected to affect the unidimensionality of the pASES-e. Since satisfaction might be an important concept to track, it's inclusion on the ASES may be appropriate as long as users recognize that it is important to consider this item as a separate construct.

The ICF does not have a core set that is specific to elbow disorders. Our current study results indicate that the core set for hand conditions can also be used for disorders of the elbow. The core set absolute linkage for the comprehensive core set is an important indicator that indicates whether a measure overall can be compared to the core set. The fact that the PREE and the ASES-e had a high score for this indicator provides initial support for the use of hand core sets to assess elbow disorders. The results of the current study indicated that the brief and comprehensive core sets for hand conditions have a wider scope with enough 'ceiling' to cover concepts of the elbow PROs. Since the core set specifically refers to other joints like shoulder, it is evident that the process for developing the core set to some extent must have considered the upper extremity is a functional unit. It is unlikely that there will be a core set developed

specifically for elbow conditions, and hence the finding that the core set for hand conditions was useful for elbow conditions supports its use in other contexts where elbow disorders are a focus. This also poses a strong case for inclusion of elbow disorders within the scope of use of the core set for hand conditions. This is of great value to clinicians who are specialised in managing upper extremity musculoskeletal disorders. Future research is needed to explore this area to achieve the goal of having one common core set for the upper extremity.

Local dependency[23] is an important issue when examining the dimensionality resulting from concept redundancy between items of the same measure. Rasch analysis is typically used to identify local dependency from a statistical perspective but we cannot explore the conceptual basis of the problem though Rasch. ICF linking might help with this. 'Measure to core set unique linkage' is a simple indicator to identify item redundancy by looking at the unique codes to which the items were linked. The difference between the core set absolute linkage and core set unique linkage indicators would give a percentage score for redundancy of concepts. Our current study results showed that there was at least a difference of 30% between the absolute and unique linkage of items of the both the PROs and both the core sets indicating redundancy of concepts between items. Redundancy in ICF can identify areas to explore, but does not necessarily mean the items are not important. For example, the redundancy is partially related to multiple pain items that all code to the ICF code but do explore different levels of pain irritability (at rest, at worst, at night, or with activity) which may be valuable in clinical conditions where pain is a predominant concern. The need for items that tap into the same ICF code can be determined statistically through factor analysis[24] and Rasch analysis.[25] Ideally, the examination of content would happen during development of measures since re-design of measures after their publication and use, create confusion and inefficiency.

Both the PREE and the ASES-e are measures for self-reported pain and disability. The 'core set unique disability representation' is a unique indicator which examines how well these measures are actually measuring the construct of disability by comparing them to a standardised framework of the ICF. Both the PREE and the ASES-e covered a wide range of disability categories on the ICF d codes ranging from d3 (communication) to d9 (Community, social and civil life). The brief core set did not have enough scope to cover all the disability items and it was saturated with a score of 100% for both the PREE and the ASES-e. This was not an

unexpected result as the purpose of a brief core set is to have the necessary codes to assess a patient condition by an individual healthcare professional and not a team.[26] Moreover both the PREE and the pASES-e are elbow specific measures with items that cover a wide range of function that may not be of great importance for an individual clinician assessment but important when a multidisciplinary team of clinicians are assessing a condition. This is supported by our finding, only 32% (PREE) and 36% (pASES-e) of the disability codes from the comprehensive core set were reflected in the content of the PROs. The purpose of the comprehensive core set is to have an exhaustive list of codes that could help assess a person's condition when a multi-disciplinary team is involved.[26] Thus this indicator validates two aspects simultaneously. Firstly it validates the use of core set for hand conditions to cover the content of elbow PROs and secondly validates the use of a PRO to measure disability.

The strengths of the current study was its standardized approach to quantifying the relationship between the PROs and the core sets while following the standardised rules to link the PROs to the ICF. The use of linkage indicators is a novel way of quantitatively analysing the depth and breadth of linkage between the measures and the core set for hand conditions and the ICF. Our study however was not without limitations. The use of two raters may have affected the selection and agreement in linking. Additional raters might have provided additional codes or insights. However, with high agreement we assume the linking was appropriately performed. However probability of this error was minimal as both the raters were formally trained on the process to link measures to the ICF and the second rater JMD was part of the expert group that approved the core set for hand conditions at the international ICF consensus conference that was conducted in Switzerland in 2009. We recommend future studies to examine linking to the hand core set of measures from different regions (i.e. shoulder) or elbow/shoulder conditions (tennis elbow or shoulder instability) to determine whether the hand core sets are useful in other contexts, and to see the usefulness of the indicators we proposed and that are published for the first time in this study.

CONCLUSION

The content of the Patient Rated Elbow Evaluation and the American Shoulder and Elbow Surgeons Elbow Questionnaire was analysed against the ICF as the standard framework and was found to align strongly with the ICF. This study has also provided preliminary evidence that the

core set for hand conditions developed for assessing conditions of the hand are well aligned with aspects of disability measured by elbow specific PRO suggesting that they may be appropriate to guide assessment and management of elbow conditions. Future studies should focus on validating the core set for hand conditions for use in assessing conditions of the whole of upper extremity.

REFERENCES

1. MacDermid JC. Outcome evaluation in patients with elbow pathology: Issues in instrument development and evaluation. *J Hand Ther*. 2001;14(2):105-114.

2. King GJ, Richards RR, Zuckerman JD, et al. A standardized method for assessment of elbow function. Research Committee, American Shoulder and Elbow Surgeons. *J Shoulder Elbow Surg*. 1999;8(4):351-354.

3. Angst F, John M, Pap G, et al. Comprehensive assessment of clinical outcome and quality of life after total elbow arthroplasty. *Arthritis Rheum*. 2005;53(1):73-82. doi: 10.1002/art.20911.

4. Angst F, Goldhahn J, Drerup S, et al. Responsiveness of five outcome measurement instruments in total elbow arthroplasty. *Arthritis Care Res (Hoboken)*. 2012. doi: 10.1002/acr.21744; 10.1002/acr.21744.

5. Vincent JI, Macdermid JC, King GJ, Grewal R. Validity and sensitivity to change of patient-reported pain and disability measures for elbow pathologies. *J Orthop Sports Phys Ther*. 2013;43(4):263-274. doi: 10.2519/jospt.2013.4029; 10.2519/jospt.2013.4029.

6. Krippendorff K. *Content analysis: An introduction to its methodology*. Sage; 2004.

7. Lynn MR. Determination and quantification of content validity. *Nurs Res*. 1986;35(6):382-385.

8. Beck CT, Gable RK. Ensuring content validity: An illustration of the process. *J Nurs Meas*. 2001;9(2):201-215.

9. World Health Organization. Towards a common language for functioning, disability and health ICF. http://www.who.int/classifications/icf/training/icfbeginnersguide.pdf. Updated 2002. Accessed 09/10, 2011.

10. Cieza A, Brockow T, Ewert T, et al. Linking health-status measurements to the international classification of functioning, disability and health. *J Rehabil Med*. 2002;34(5):205-210.

11. Cieza A, Geyh S, Chatterji S, Kostanjsek N, Ustun B, Stucki G. ICF linking rules: An update based on lessons learned. *J Rehabil Med*. 2005;37(4):212-218. doi: 10.1080/16501970510040263.

12. Cieza A, Stucki G. Content comparison of health-related quality of life (HRQOL) instruments based on the international classification of functioning, disability and health (ICF). *Qual Life Res*. 2005;14(5):1225-1237.

13. Liem IS, Kolling C, Marks M, Nelissen RG, Goldhahn J. Development of a score set to measure function and quality of life in patients suffering from elbow pathology. *Arch Orthop Trauma Surg*. 2012;132(6):831-837. doi: 10.1007/s00402-012-1472-0.

14. Simmen BR, Angst F, Schwyzer HK, et al. A concept for comprehensively measuring health, function and quality of life following orthopaedic interventions of the upper extremity. *Arch Orthop Trauma Surg*. 2009;129(1):113-118. doi: 10.1007/s00402-008-0718-3.

15. John M, Angst F, Awiszus F, King GJ, MacDermid JC, Simmen BR. The american shoulder and elbow surgeons elbow questionnaire: Cross-cultural adaptation into german and evaluation of its psychometric properties. *J Hand Ther*. 2010;23(3):301-13.

16. John M, Angst F, Pap G, Junge A, Mannion AF. Cross-cultural adaptation, reliability and validity of the patient rated elbow evaluation (PREE) for german-speaking patients. *Clin Exp Rheumatol*. 2007;25(2):195-205.

17. Hanyu T, Watanabe M, Masatomi T, et al. Reliability, validity, and responsiveness of the japanese version of the patient-rated elbow evaluation. *Journal of Orthopaedic Science*. 2013;18(5):712-719. Accessed 21 January 2014.

18. ICF Research Branch. ICF core set for hand conditions. http://www.icf-research-branch.org/icf-core-sets-projects-sp-1641024398/other-health-conditions/development-of-icf-core-sets-for-hand-conditions. Updated 2009. Accessed 02/24, 2012.

19. MacDermid JC. The patient rated elbow evaluation (PREE) - user manual. http://www.srs-mcmaster.ca/Portals/20/pdf/research_resources/PREE%20User%20Manual_June%202010.pdf. Published June, 2010. Updated June, 2010. Accessed 08/05, 2012.

20. Grayson K, Rust R. Interrater reliability. *Journal of Consumer Psychology*. 2001;10(1-2):71-73.

21. Arumugam V, Macdermid JC, Grewal R. Content analysis of work limitation, stanford presenteeism, and work instability questionnaires using international classification of functioning, disability, and health and item perspective framework. *Rehabil Res Pract*. 2013;2013:614825.

22. MacDermid JC. International Classification of Functioning Disability and Health (ICF) Linkage Indicators. http://www.srs-mcmaster.ca/Portals/20/pdf/research_resources/ICFlinkageindicators.pdf. Updated 2014. Accessed 06/05, 2014.

23. Fischer GH, Molenaar IW. *Rasch models: Foundations, recent developments, and applications.* New York: Springer-Verlag; 1995:436.

24. Harman HH. *Modern factor analysis.* Oxford, England: Univ. of Chicago Press; 1960:xvi, 469.

25. Andrich D. *Rasch models for measurement.* Vol 07-068. Newbury Park: Sage Publications; 1988:95.

26. Selb M, Escorpizo R, Kostanjsek N, Stucki G, Ustun B, Cieza A. A guide on how to develop an international classification of functioning, disability and health core set. *Eur J Phys Rehabil Med*. 2014.

Table 1: ICF linkage indicator for measure to ICF linkage and percentage agreement between the raters

	PREE (%)	ASES-e (%)
Measure to ICF linkage	100	94
Percentage agreement between raters	96	95

PREE - The Patient Rated elbow evaluation; ASES-e - The self-report section of the American Shoulder and Elbow Surgeons society – Elbow form

Table 2: ICF linkage indicators to define the depth and breadth of linkage between items of the PREE and pASES-e and the ICF core sets

	Linkage indicator	PREE		pASES-e	
		Brief core set	**Comprehensive core set**	**Brief core set**	**Comprehensive core set**
1	Measure to Core Set Absolute Linkage	85%	100%	84%	95%
2	Measure to Core Set Unique Linkage	50%	65%	58%	68%
3	Core set representation	74%	17%	70%	15%
4	Unique core set disability representation	100%	32%	100%	26%

PREE - The Patient Rated elbow evaluation; ASES-e - The self-report section of the American Shoulder and Elbow Surgeons society – Elbow form

Table 3: ICF linking of Patient Rated Elbow Evaluation

S.NO	ITEM	CODE	DESCRIPTION
		Pain	
1.	When it is at its worst	b280,	Sensation of pain
2.	At rest	b280	Sensation of pain
3.	When lifting a heavy object	b280,	Sensation of pain,
		d430	Lifting and carrying objects
4.	When doing a task with repeated elbow	b280,	Sensation of pain
	movement	b710	Mobility of joint functions
5.	How often do you have pain	b280,	Sensation of pain
		b710	Mobility of joint functions
		Specific Activities	
6.	Comb my hair	d520	Caring for hair
7.	Eat with fork or spoon	d550	Eating
8.	Pull a heavy object	d445	Hand and arm use
9.	Use my arm to rise from a chair	d410	Changing basic body position
10.	Carry a 10lb object with my arm at my side	d430	Lifting and carrying objects
11.	Throw a small object, such as a tennis ball	d445	Hand and arm use
12.	Use a telephone	d360	Using communication devices
13.	Do up buttons on the front of my shirt	d540	Dressing
14.	Wash my opposite armpit	d510	Washing oneself
15.	Tie my shoe	d540	Dressing
16.	Turn the doorknob and open a door	d445	Hand and arm use
		Usual Activities	
17.	Personal activities (dressing, washing)	d540,	Dressing
		d510	Washing oneself
18.	Household work (cleaning, maintenance)	d630 to	Household tasks
		d649	
19.	Work (your job or everyday work)	d840 to	Work and employment
		d859	
20.	Recreational activities	d920	Recreation and leisure

Table 4: ICF linking of American Shoulder and Elbow Surgeons – Elbow form

S.NO	ITEM	CODE	DESCRIPTION
		Pain	
1.	When it is at its worst	b280	Sensation of pain
2.	At rest	b280	Sensation of pain
3.	Lifting a heavy object	b280, d430	Sensation of pain, Lifting and carrying objects
4.	When doing a task with repeated elbow movements	b280, b710	Sensation of pain, Mobility of joint functions
5.	At night	b280	Sensation of pain
		Function	
6.	Do up top button on shirt	d540	Dressing
7.	Manage toileting	d530	Toileting
8.	Comb hair	d520	Carrying for hair
9.	The shoes	d540	Dressing
10.	Eat with utensil	d550	Eating
11.	Carry a heavy object	d430	Lifting and carrying objects
12.	Rise from chair pushing with arm	d410	Changing and maintaining body function
13.	Do heavy household chores	d649	Household tasks
14.	Turn a key	d445	Hand and arm use
15.	Throw a ball	d445	Hand and arm use
16.	Do usual work	d840 to d859	Work and employment
17.	Do usual sport	d920	Recreation and leisure
		Satisfaction	
18.	How satisfied are you with the surgery?	nc	Not covered by ICF

nc – not codable

Figure 1: ICF linkage indicator for measure to ICF linkage and percentage agreement between the raters

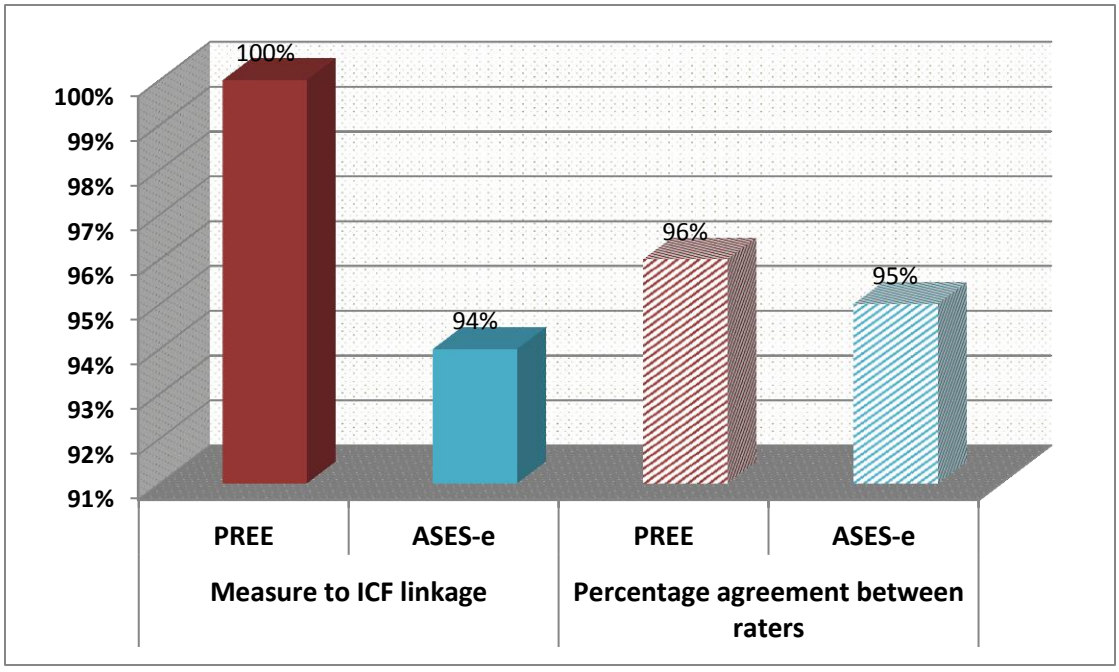
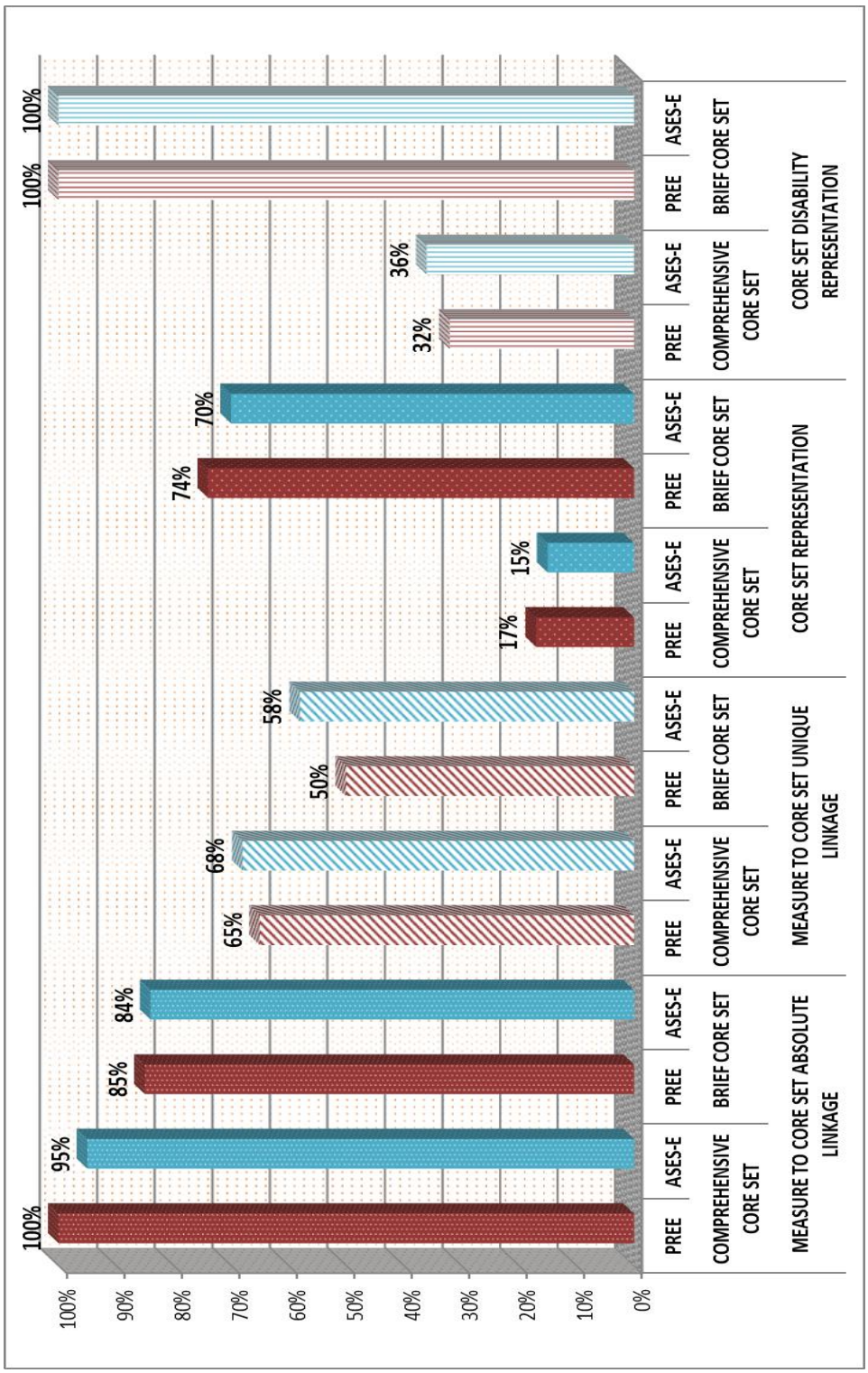


Figure 2: ICF linkage indicators to define the depth and breadth of linkage between items of the PREE and ASES-e and the ICF core sets



CHAPTER 5: THE PATIENT RATED ELBOW EVALUATION AND THE AMERICAN SHOULDER AND ELBOW SURGEONS – ELBOW FORM CAPTURE ASPECTS OF FUNCTIONING THAT ARE IMPORTANT TO PATIENTS WITH ELBOW INJURIES

ABSTRACT

Background and purpose: The Patient Rated Elbow Evaluation (PREE) and the self-report section of the American Shoulder Elbow Surgeons – elbow form (pASES-e) are two important elbow specific self-report measures used in routine clinical practice. The objective of the current study is to use the International Classification of Functioning Disability and Health (ICF) to link aspects of functioning that are reported using the Patient Specific Functional Scale (PSFS) by a cohort of patients with elbow disorders and compare it to the content of the PREE and the pASES-e.

Design: Cross sectional study

Methods: One hundred patients with a variety of elbow disorders (Mean age and SD 53.88 (14.51); Range 16 to 89; M: F 48: 52) were recruited from the Roth McFarlane Hand and Upper Limb Center. They self-endorsed important aspects of functioning using the PSFS. These concerns were linked to the ICF using formal linking procedures. The linked ICF categories were compared to those obtained by linking the items of the PREE and the pASES-e. Linking was carried out by two independent raters and agreement was calculated using Kappa.

Results: A total of 423 self-reported functional activities were linked to 25 second level ICF categories from the activity and participation domain. D640 doing housework was the most common code (52%) followed by D540 Dressing (47%) and D475 Driving (35%). 71% and 50% of the ICF categories used to link self-endorsed function were captured by the PREE and the pASES-e respectively. D475 –driving (35%) and D440 - Fine hand use (24%) were the 2 major categories that were not captured by the questionnaires.

Limitations: Patients were from a tertiary center and may not reflect the concerns of patient attending a general practice.

Conclusion: The PREE provided more comprehensive coverage of patients' functional concerns. Important aspects of function may be missed by sole reliance on standard self-report measures.

Keywords: ICF; PSFS; PREE; pASES-e; Elbow disorders

INTRODUCTION

Healthcare has evolved to place greater emphasis on patient empowerment in the management of their personal health. Patient empowerment can achieve better patient participation and self-management of conditions.[1, 2] The concept of patient empowerment is a multi-level construct. One means of promoting greater patient empowerment is including patients in the assessment of their state of health or functioning in addition to the traditional objective assessments performed by their health care provider.[2, 3] Patient reported outcome measures (PROM) have been increasingly used in practice.[4, 5] Previous studies have highlighted that PROMs complement traditional clinician based outcome measures (CBO), allowing a holistic view of the health status of patients.[3, 6]

Elbow pain and the resulting functional limitation can be caused by various structures in dysfunction resulting in tendinitis, strains, sprains, bursitis, arthritis fractures, dislocations, etc. depending on the anatomical structure that is involved. Although treatment varies by condition,[7] common functional problems can be expected with elbow pathology. The elbow is the vital link between the shoulder and the hand, helping to place the hand at target positions that would allow it to perform various fine and gross motor activities.[8] There are PROMs that are available to measure functional limitation and pain caused by disorders of the elbow.[9] Two of the commonly used PROMs are the Patient Rated Elbow Evaluation (PREE)[10, 11] and the self-report section of the American Shoulder Elbow Surgeons – elbow form (pASES-e).[12] Both these measures were previously validated and have been reported to exhibit excellent psychometric properties.[10, 13-18]

The International Classification of Functioning Disability and Health (ICF) is a biopsychosocial model providing a standard framework and classification to describe health and health-related states.[19] The ICF consists of the model and a hierarchical coding system to describe health and functioning. The ICF has three domains: body functions and structure; activity and participation and contextual factors, which include environmental factors and personal factors. The coding system contains categories that are placed under each of the domains. Personal factors do not have categories.[19, 20]

The ICF model has been previously used to create a core set of measures that would help measuring quality of life, objective and subjective function in the upper extremity and also in

elbow pathologies.[16, 21] Simmen et al[21] concluded that for assessment of the elbow along with the other measures the PREE or the pASES-e can be used as elbow-specific patient reported questionnaire to measure elbow function, while Liem et al[16] have recommended the use of PREE for the same purpose. Both of these studies have completed an initial validation of their proposed measures and have recommended the use of them. Recently, these two PROMs have undergone formal content comparison using the ICF linking process as described in the literature.[22] It was found that the PREE and the pASES-e align with the ICF framework and the subset of codes contained on the core set for Hand Conditions. [22] This study concluded that both measures had enough breadth and depth to help assess function in disorders of the elbow when considering the ICF as a criterion comparator.[22] However the above mentioned studies did not address the extent to which these items reflected patient concerns.

As the PREE and the pASES-e have been recommended for use in the score set to assess elbow disorders, they should also be reflective of the concept of patient centered care. As there is limited information about the extent to which patients are instrumental in generation of items for both the PREE and the pASES-e, it is important to examine the extent to which they reflect patient concerns. Combining this information with existing evidence about the clinical measurement properties, will provide confidence in patient-centered, evidence-based evaluation of outcomes of patients with elbow disorders.[24] The ICF is a standard for classification of health related states and is an ideal criterion for content analysis and validation in functional outcome measures.[25]

The objective of the current study is to use the ICF classification to summarize and identify aspects of functioning that are reported as important by a cohort of patients with elbow disorders and compare it to the content of the PREE and the pASES-e.

METHODS

Subjects: One hundred patients (Mean age and SD 53.88 (14.51); Range 16 to 89; M: F 48: 52) with a variety of elbow disorders were recruited from the Roth McFarlane Hand and Upper Limb Center at St. Joseph's healthcare London, London ON, Canada. (See Table 1) Once they agreed to participate, subjects signed an informed consent form. (Appendix 2&3)

Procedure: To identify the aspects of functioning that are important to patients with elbow disorders the Patient Specific Functioning Scale (PSFS)[26] was used to collect self-nominated functional concerns. We used this as previous reports have found that items generated through the PSFS are more representative of the activity and participation domain of the ICF.[27] Participants were asked to fill the questionnaire at the clinic waiting room. The PSFS is a self-report measure allowing patients to identify up to 5 functional limitations and rate them on a scale from 0 to 10, with 0 being "Unable to perform activity " and 10 being "Able to perform activity at the same level as before injury or problem". This questionnaire has been validated in many different conditions like knee dysfunction,[28] radiculopathy,[29] neck dysfunction,[30] etc. It was found to be valid, reliable and responsive.[28-30] As the PSFS was used for item generation, all nominated items were included in the analysis.

ICF linking of patient concerns: Once the data was collected, all the concerns were linked to the ICF by a pair of raters independently. Both the raters had experience with ICF linking. The first rater had more experience was part of the committee that contributed to the ICF Core Set for Hand Conditions and trained the second rater [23] ICF linking rules proposed by Cieza and colleagues were used.[24, 31, 32] Each of the responses was linked with alphanumeric categories to the most specific level that is possible. The ICF coding book[33] and the ICF browser[33, 34] were used during linking. Only the first two hierarchical levels of categories were considered for analysis to give a better overview of the areas that were highlighted by the responses to the PSFS. If two items listed by a patient were linked to the same code, then that code was counted only one time for that person. Once coding was complete, reviewers met to discuss the categories and discuss discrepancies to achieve consensus. A third rater was available to resolve any persistent disagreement.

Comparison to the PREE and the pASES-e: Once patient concerns were linked to the ICF, the categories used were compared to those used to link the PREE and the pASES-e. The codes represented on the PREE and the pASES-e have been described in the previous study.[22]

Data analysis

All statistical analysis was performed using SPSS version 22. Agreement was assessed by calculating Kappa.[35, 36] Frequencies of each of the categories were also calculated. It was agreed

upon that a higher frequency in categories appearing while linking patient concerns indicates greater importance. If a category was nominated by more than 25% of the subjects, the functional concern was considered of high importance, while anything between >10% and ≤ 25% indicate concerns of moderate importance. Categories that were used on 10% or less of the subjects were considered as less important to our cohort of patients; however we recognize that even infrequently reported codes are of concern to a subset of patients. These arbitrary benchmarks were assigned to allow us to interpret our findings.

RESULTS

Linking of patient concerns to the ICF: One hundred participants nominated 469 functional activities which they felt were important and affected by their elbow condition. All the concerns were linked to the ICF. If an ICF category appeared twice for an individual then the duplicates were removed. Once duplicate ICF categories were removed, 423 concerns remained which were then analyzed. (See table 2) All the categories used to link these concerns to the ICF were from the activity and participation domain ('d' categories). Twenty five unique second level categories were used to link the concerns, from chapters D4 (Mobility) to D9 (Community, social and civic life). Generally we did not have any problems with the linking with exception of one problem that we encountered with the patient concern 'babysitting'; this can be linked to two ICF categories D 855 Non remunerative employment and D 7603 Extended family relationships, but ultimately it was decided that non-remunerative employment would be more appropriate . Agreement between the raters was very high (Kappa =100%).

Based on the predefined benchmarks that were proposed to identify the level of importance of individual categories, 11 of the ICF categories that were used for less than 10% of the participants were removed resulting in 14 second level ICF categories that were used for final analysis. The category with the highest frequency was D640 – Doing housework, it was linked to concerns of 52% of the participants. Since our sample was gender-balanced, this included a variety of traditionally masculine or feminine and more neutral roles such as cleaning the house, shoveling snow, gardening etc. (See Table 2) On the whole, ten categories were classified as highly important to our cohort of patients (>25 %); this included D 540 Dressing; D 920 Recreation and leisure; D 475 Driving; D 440 Fine hand use amongst others. Four categories were of moderate importance (10 – 25 %) to our cohort. The categories that were included in this

tier were D 445 Hand and arm use; D 520 Caring for body parts; D 570 Looking after one's health and D 650 Caring for household objects. (See Figure 1) While categories like D 410 changing body position; D 530 Toileting; D 750 Informal social relationships were less frequently nominated as being important functional concerns(<10%). (See Table 2)

Comparison to the PREE and the pASES-e: The extent to which, the nominated aspects of functioning were captured by the two PROMs was different. (See Table 2) The PREE was able to capture almost three fourths (10 out of the 14; 71%) of the ICF categories identified as important by the participants of the study while the pASES-e captured only half (7 out of 14 ICF categories; 50%). This analysis also pointed out some aspects of functioning that were important to our participants but were not captured either by the PREE or the pASES-e; they are D475 - Driving (35%); D440 - Fine hand use (24%); D570 – Looking after one's health (17%) and D650 – caring for household objects (11%). In addition, the pASES-e did not capture 3 important categories; of which D640 – Doing housework (52%) which was identified as the most important category was one. There were also a few categories that were captured by the PREE and the pASES-e that were less frequently identified as important by our participants; they are D410 – Changing basic body position (3%); D530 - toileting (2%); D855 – Non-remunerative employment (2%); and D360 – Using telecommunication devices and techniques (1%).

DISCUSSION

This study established that the PREE and the pASES-e were able to capture aspects of functioning important to patients and that align with the ICF; with this happening to a greater extent on the PREE than the ASES-e. Since all patients reported concerns were from the activity and participation section ('d' categories) of ICF, this validated that the PROM measure this conceptual domain. This was expected as the PSFS was designed to cover the activity and participation domain of the ICF.[27, 19]

The exercise of using ICF language to compare the content of the elbow PROMs to the aspects of functioning with that reported, without cueing, by patients has two implications. First, it suggests that these PROMs may be useful in collaborative goal setting since they cover patient-important aspects of functioning. Greater patient participation in setting goals for treatment can contribute to patient empowerment[3] and may facilitate adherence to treatment.

This also suggests that one of the reasons patients value their clinicians using a PROM is that they find them relevant.

Another contribution of this work is providing content validation data on the PREE and the pASES-e, as a PROM for elbow conditions, since these were not previously subjected to a formal test of their content validity.[10, 12] Hence, this study with linking the PREE and pASES-e to the ICF supports the content validity against the current standard in measurement of functioning and disability, the ICF.[25] The PREE and the pASES-e were developed independently at different points in time. We found that 64% of the ICF categories appeared on both PROMs. Linking to the same code adds to the construct validity of these two PROMs and also clearly indicates that experts agreed on the importance of these concepts.

The breadth of item coverage indicates the bandwidth of the instrument while the depth of coverage within an area of content indicates the precision of the instrument.[24] The PREE and the pASES-e have demonstrated acceptable breadth to capture diverse in the aspects of functioning identified by the patients. Patient concerns were spread across from D4 (Mobility) to D9 (Community, social and civic life), while the ICF categories used to link the PREE were from D3 (Communication) to D9 (Community, social and civic life) and for the pASES-e they were from D4 (Mobility) to D9 (Community, social and civic life). This adds evidence to the acceptable level of band-width demonstrated by the 2 PROMs; and their ability to acknowledge and identify the diversity of the difficulties experienced by patients with elbow disorders. However, when comparing the depth of the PROMs, the PREE was better, capturing almost three fourths of the ICF categories linked to the patient data. This suggests that the PREE provides better coverage of function than the pASES-e; which may explain findings of previous psychometric studies that have made similar conclusions about performance of these 2 PROM based on quantitative analysis.[13, 14]

One major aspect of functioning that was totally missed by both the PROMs was driving. Driving has become an important aspect of everyday living, being routinely used by a substantial subset of people and a greater necessity where public transport is absent. For example in the UK, in 2012, trips made by car were identified as 68% of all trips.[37] Previous studies have identified driving as an important component that affects the ability of a patient to function independently.[38-40] This concept is infrequently covered by PROMs specific to other regions of

the body. A previous study on patients with wrist injuries has identified a similar trend; they found that none of the 8 PROMs measuring pain and disability in wrist conditions covered the concept of driving.[41] Conversely, the Neck Disability Index, does contain driving as an item and it has been one of the most problematic items on the measure, and is frequently left blank by participants who do not drive.[42] In some countries, this has been extremely problematic since females are less likely to drive in some cultures. The implications of this are that clinicians may need to consider driving and transportation needs as an issue to explore with patients, which will not be detected by routine PROM administration. Future revisions to PROM that wish to incorporate the concept, may need to use more global items like "transportation", or "driving/being a passenger" to be sufficiently inclusive since missing items are a problem in outcome measurement.

'Fine hand use' was the other important concern that the 2 PROMs missed to capture. Generally, the primary role in fine hand use goes to the hand to perform specific fine motor tasks. A previous study by Coehnon and colleagues[41] identified that all the 8 PROMs measuring hand function included fine hand use as one of their items. The elbow couples with shoulder to position the hand in space to perform fine motor activities.[8] While it is clear that fine hand use is function of the hand, it sparks a debate whether this should be included in elbow outcome measures with the elbow playing an indirect role. The limiting factor in the hand tasks cited by patients was the ability to place the hand in the correct position due to limited elbow motion. Some elbow questionnaires use the item "using a telephone" to capture the elbow motion needed to bring a functional item into range. However, more modern cell phones are often not used with the same elbow range of motion requirements as older land-based telephones. Thus, over time the content validity of this item may have deteriorated. Qualitative interviews are needed to find ways to generate an item that would capture difficulty positioning/stabilizing the hand in order to accomplish fine hand activity - to maintain patient centeredness and theoretical construct validity.

Exercising was another functional activity that was missed in the content of both the PROMs. More and more people are embracing exercises to maintain optimal health. According to a report from World health Organization (WHO) on physical activity globally 69% of adults aged above 15 were active.[43] This means more than two thirds of the world's population does

some kind of physical activity regularly. Many forms of physical activity or exercise would involve elbow motion, strength and stability. Ability to maintain personal health was also cited by patients with distal radius fractures as a concern in a similar study.[44] This justifies the inclusion of exercises in to these PROMs in the future.

The strengths of the current study are sampling a substantial number of patients with a variety of elbow conditions. The ICF is considered the gold standard in measurement of functioning and disability[25] and provided a strong reference standard. Our study also had some limitations; since participants reported the important aspects of functioning in a structured written format, there was very minimal opportunity to check the intended meaning with the patient. Qualitative interviews with smaller numbers of patients may provide a deeper understanding of patients' functional concerns. However, qualitative approaches cannot assume generalizability. We mitigated the potential for misclassification, by checking over items generated by patients and clarifying where necessary, while patients were in attendance. Secondly, our patients were from a tertiary care center which might prevent the generalizability of the results of this study to general population. However, there is no rationale for why the functional concerns of patients to attend a tertiary center would be different than the general public from which they are drawn.

CONCLUSION

The current study identified aspects of functioning that were important to patients with elbow disorders and was able to identify a diverse list of functional activities involving the elbow. The use of ICF to compare the content of the PREE and the pASES-e to the content of patient data indicated that the PREE had demonstrated optimal levels of breadth and depth to capture majority of the aspects of functioning identified as important by patients. Future studies should incorporate items like driving and exercise into these PROMs to improve their scope and precision. Future studies might explore how driving/transportation needs and positioning the hand for fine hand use might be optimally addressed. PROMs should be revisited at least every 10 years to adapt to the changing trends in the society and culture which might deem some of the items unfit while pointing out to the need for new contemporary items.

REFERENCES

1. Mavis B, Holmes Rovner M, Jorgenson S, et al. Patient participation in clinical encounters: A systematic review to identify self-report measures. *Health Expectations*. 2014: 10.1111/hex.12186.

2. Haywood K, Marshall S, Fitzpatrick R. Patient participation in the consultation process: A structured review of intervention strategies. *Patient Educ Couns*. 2006;63(1-2):12-23.

3. Robinson CA, Shumway-Cook A, Ciol MA, Kartin D. Participation in community walking following stroke: Subjective versus objective measures and the impact of personal factors. *Phys Ther*. 2011;91(12):1865-1876.

4. Food and Drug Administration.Patient reported outcome measures: Use in medical product development to support labeling claims. . 2009;2011(11/8/2011). Accessed 11/8/2011.

5. MacDermid JC, Stratford P. Applying evidence on outcome measures to hand therapy practice. *Journal of Hand Therapy*. 2004;17(2):165-73. 6. Slevin ML, Plant H, Lynch D, Drinkwater J, Gregory WM. Who should measure quality of life, the doctor or the patient? *Br J Cancer*. 1988;57(1):109-112.

7. Morrey BF. The elbow and its disorders, 4th ed. *Scitech Book News*. 2009;33(2):n/a.

8. Lockard M. Clinical biomechanics of the elbow. *Journal of Hand Therapy*. 2006;19(2):72-81.

9. Longo UG, Franceschi F, Loppini M, Maffulli N, Denaro V. Rating systems for evaluation of the elbow. *Br Med Bull*. 2008;87(1):131-161.

10. MacDermid JC. Outcome evaluation in patients with elbow pathology: Issues in instrument development and evaluation. *Journal of Hand Therapy*. 2001;14(2):105-114. 11. MacDermid JC. Patient-Rated Elbow Evaluation. . http://www.srs-mcmaster.ca/Portals/20/pdf/research_resources/PREE.pdf. Accessed 09/10, 2011.

12. King GJ, Richards RR, Zuckerman JD, et al. A standardized method for assessment of elbow function. research committee, american shoulder and elbow surgeons. *J Shoulder Elbow Surg*. 1999;8(4):351-354.

13. Vincent JI, King GJW, Grewal R. Validity and sensitivity to change of patient-reported pain and disability measures for elbow pathologies. *J Orthop Sports Phys Ther*. 2013;43(4):263-274.

14. Angst F, Goldhahn J, Drerup S, et al. Responsiveness of five outcome measurement instruments in total elbow arthroplasty. *Arthritis Care and Research*. 2012;64(11):1749-1755.

15. Angst F, John M, Pap G, et al. Comprehensive assessment of clinical outcome and quality of life after total elbow arthroplasty. *Arthritis Care and Research*. 2005;53(1):73-82.

16. Liem IS, Kolling C, Marks M, Nelissen RG, Goldhahn J. Development of a score set to measure function and quality of life in patients suffering from elbow pathology. *Arch Orthop Trauma Surg*. 2012;132(6):831-837.

17. Vincent J, MacDermid JC. The patient-rated elbow evaluation (PREE). *Journal of physiotherapy*. 2012;58(4):274.

18. Vincent JI, Macdermid JC, King GJ, Grewal R, Lalone E. A systematic review of the psychometric properties of two elbow self-report pain and disability measures. *Arthritis Care & Research*. 2014;Submitted.

19. Cieza A, Stucki G. The international classification of functioning disability and health: Its development process and content validity. *Eur J Phys Rehabil Med*. 2008;44(3):303-313.

20. World Health Organization. Towards a Common Language for Functioning, Disability and Health ICF. . http://www.who.int/classifications/icf/training/icfbeginnersguide.pdf. Updated 2002. Accessed 09/10, 2011.

21. Simmen BR, Angst F, Schwyzer HK, et al. A concept for comprehensively measuring health, function and quality of life following orthopaedic interventions of the upper extremity. *Arch Orthop Trauma Surg*. 2009;129(1):113-118.

22. Vincent JI, Macdermid JC, King GJ, Grewal R. Linking of the patient rated elbow evaluation (PREE) and the american shoulder and elbow surgeons – elbow questionnaire (pASES-e) to the international classification of functioning disability and health (ICF) and hand core sets . *Journal of Hand Therapy*. 2014;Submitted.

23. Rudolf K, Kus S, Chung KC, Johnston M, LeBlanc M, Cieza A. Development of the international classification of functioning, disability and health core sets for hand conditions results of the world health organization international consensus process. *Disabil Rehabil*. 2012;34(8):681-693.

24. Cieza A, Stucki G. Content comparison of health-related quality of life (HRQOL) instruments based on the international classification of functioning, disability and health (ICF). *Quality of Life Research*. 2005;14(5):1225-1237.

25. Weigl M, Cieza A, Harder M, et al. Linking osteoarthritis-specific health-status measures to the international classification of functioning, disability, and health (ICF). *Osteoarthritis and Cartilage*. 2003;11(7):519-523.

26. Stratford P. Assessing disability and change on individual patients: A report of a patient specific measure. *Physiotherapy Canada*. 1995;47(4):258-263.

27. Fairbairn K, May K, Yang Y, Balasundar S, Hefford C, Abbott JH. Mapping patient-specific functional scale (PSFS) items to the international classification of functioning, disability and health (ICF). *Phys Ther*. 2012;92(2):310-317.

28. Chatman AB, Hyams SP, Neel JM, et al. The patient-specific functional scale: Measurement properties in patients with knee dysfunction. *Phys Ther*. 1997;77(8):820-829.

29. Young IA, Cleland JA, Michener LA, Brown C. Reliability, construct validity, and responsiveness of the neck disability index, patient-specific functional scale, and numeric pain rating scale in patients with cervical radiculopathy. *Am J Phys Med Rehabil*. 2010;89(10):831-839.

30. Westaway MD, Stratford PW, Binkley JM. The patient-specific functional scale: Validation of its use in persons with neck dysfunction. *J Orthop Sports Phys Ther*. 1998;27(5):331-338.

31. Cieza A, Geyh S, Chatterji S, Kostanjsek N, Ustun B, Stucki G. ICF linking rules: An update based on lessons learned. *J Rehabil Med*. 2005;37(4):212-218.

32. Cieza A, Brockow T, Ewert T, et al. Linking health-status measurements to the international classification of functioning, disability and health. *J Rehabil Med*. 2002;34(5):205-210.

33. Organization WH. *International Classification of Functioning, Disability and Health: ICF*. World Health Organization; 2001.

34. World Health Organisation. International Classification of Functioning, Disability and Health (ICF) Browser. . http://apps.who.int/classifications/icfbrowser/. Updated 2013. Accessed 04/14, 2014.

35. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.

36. Sim J, Wright CC. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Phys Ther*. 2005;85(3):257-268.

37. Department for Transport (United Kingdom). National Travel Survey: 2012. . https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/243957/nts2012-01.pdf. Updated 2013. Accessed 06/26, 2014.

38. Busteed S, Daly M, Silke C, Molloy MG. Rheumatoid arthritis impairs driving ability even in patients with a low disability index. *Rheumatology*. 2004;43(1):107-108.

39. Cranney AB, Harrison A, Ruhland L, et al. Driving problems in patients with rheumatoid arthritis. *J Rheumatol*. 2005;32(12):2337-2342.

40. Fan A, Wilson KG, Acharya M, Cranney A, Buenger U, Marshall S. Self-reported issues with driving in patients with chronic pain. *PM R*. 2012;4(2):87-95.

41. Coenen M, Kus S, Rudolf K, et al. Do patient-reported outcome measures capture functioning aspects and environmental factors important to individuals with injuries or disorders of the hand? *Journal of Hand Therapy*. 2013;26(4):332.

42. MacDermid JC, Walton DM, Avery S, et al. Measurement properties of the neck disability index: A systematic review. *J Orthop Sports Phys Ther*. 2009;39(5):400-417.

43. World Health Organization (WHO). Physical Inactivity: A Global Public Health Problem.http://www.who.int/dietphysicalactivity/factsheet_inactivity/en/. Updated 2014. Accessed 06/26, 2014.

44. MacDermid JC, Kus S, Beavis D, et al. ICF linking of participation restrictions reported by a cohort of 403 patients with a distal radius fracture. (submitted).Journal of Hand therapy. 2014.

Table 1: Patient characteristics

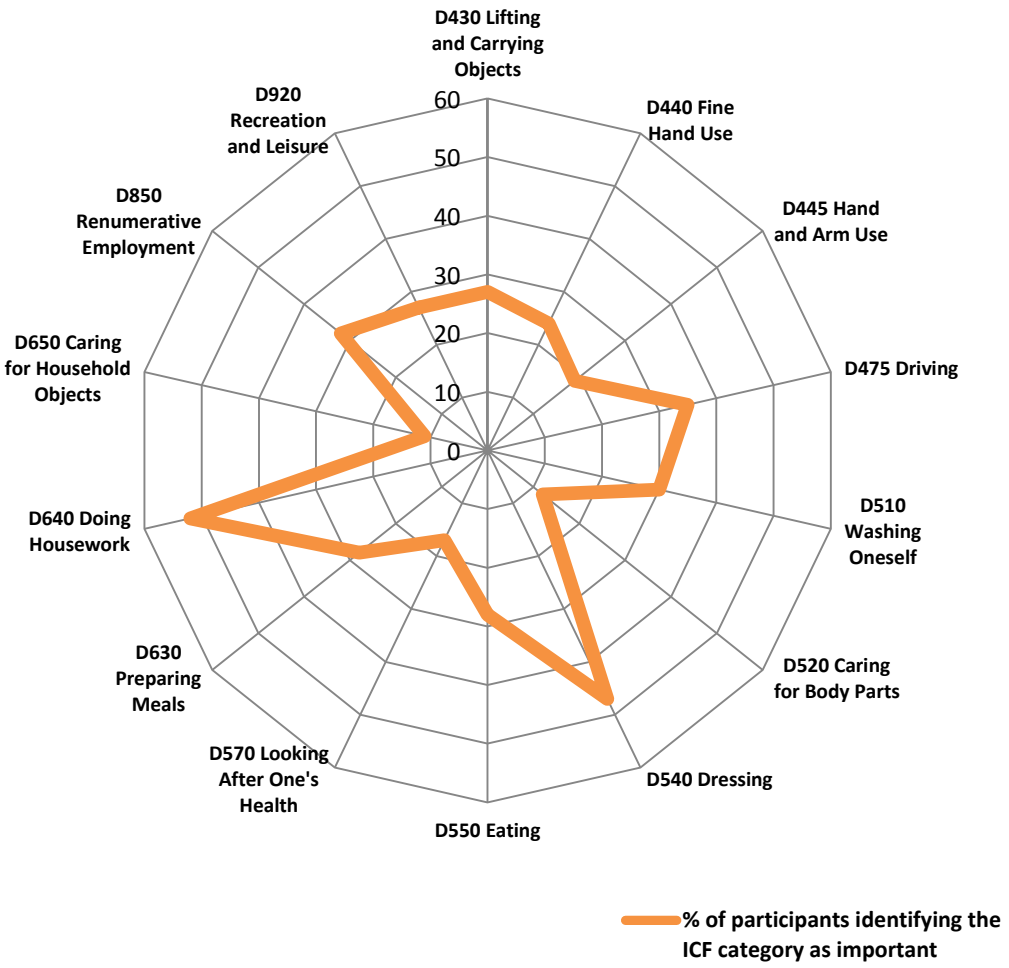
Characteristics	Values
	n = 100
Male : Female	48:52
Age (mean and SD)	53.88 (14.51)
Range	19 to 89 years
Surgery (Yes: No)	85 : 15
	Diagnosis %
Osteoarthritis	4
Tennis elbow	1
Bursitis	1
Elbow pain	2
Elbow contractures / stiffness	3
Ulnar neuritis	7
Ligament tear	1
Biceps rupture	9
Triceps rupture	3
Terrible triad	1
Distal humerus fractures	19
Radial head fractures	37
Proximal ulnar fractures	11
Failed total elbow arthroplasty	1

Table 2: Table showing the aspects of functioning captured by PREE and pASES-e

S. No	ICF code	%	Description	PREE	pASES-e
1	D170	6	Writing		
2	D210	1	Undertaking a single task		
3	D360	1	Using communication devices and techniques	☑	
4	D410	3	Changing basic body position	☑	☑
5	D415	7	Maintaining a body position		
6	D420	2	Transferring oneself		
7	D430	27	Lifting and carrying objects	☑	☑
8	D440	24	Fine hand use		
9	D445	19	Hand and arm use	☑	☑
10	D450	1	Walking		
11	D455	5	Moving around		
12	D475	35	Driving		
13	D510	30	Washing oneself	☑	
14	D520	12	Caring for body parts	☑	☑
15	D530	2	Toileting		☑
16	D540	47	Dressing	☑	☑
17	D550	28	Eating	☑	☑
18	D570	17	Looking after one's health		
19	D630	28	Preparing meals	☑	
20	D640	52	Doing housework	☑	
21	D650	11	Caring for household objects		
22	D760	3	Family relationships		
23	D850	32	Remunerative employment	☑	☑
24	D855	2	Non-remunerative employment	☑	☑
25	D920	27	Recreation and leisure	☑	☑

PREE - Patient Rated Elbow Evaluation; pASES-e - self-report section of the American Shoulder Elbow Surgeons – elbow form; % - percentage of participants reported the ICF category as important

Figure 1: Spider plot showing the distribution ICF categories and their importance based on patient responses



CHAPTER 6: RASCH ANALYSIS OF THE PATIENT RATED ELBOW EVALUATION

ABSTRACT

Background: The Patient Rated Elbow Evaluation (PREE) was developed as an elbow joint specific measure of pain and disability and validated with classical psychometric methods. More recently, Rasch analysis has contributed new methods for analyzing the clinical measurement properties of self-report outcome measures. The objective of the study was to determine aspects of validity of the PREE using the Rasch model to assess the overall fit of the PREE data, the response scaling, individual item fit, differential item functioning (DIF), local dependency, unidimensionality and person separation index (PSI).

Methods: A convenience sample of 236 patients (Age range 21-79 years; M: F- 97:139) with elbow disorders where recruited from the Roth McFarlane Hand and Upper Limb Centre, London, Ontario, Canada. Patients completed the PREE at baseline after surgery and this data set was used for analysis. Rasch analysis was conducted using RUMM 2030 software on the 3 sub scales of the PREE separately.

Results: The 3 sub scales showed misfit initially with disordered thresholds (17 out of 20 items), uniform DIF with two items ("Carrying a 10lbs object" from specific activities subscale; and "household work" from the usual activities subscale); multidimensionality and local dependency. The Pain subscale satisfied Rasch expectations when item 2 "Pain – At rest" was split for age group, while the usual activities subscale readily stood up to Rasch requirements when the item 2 "household work" was split for gender. The specific activities sub scale demonstrated fit to the Rasch model when sub test analysis was performed to cancel out local dependency. All three subscales of the PREE were well targeted and had high reliability (PSI >0.80).

Conclusion: This study has indicated that the three sub scales of the PREE are reliable and can provide interval level scaling with rescoring of the 0-10 scale. In future clinical applications and research studies, these alternative scoring procedures can be used to obtain an unbiased linear interval level data for the PREE.

Keywords: Patient Rated Elbow Evaluation, Rasch analysis, Elbow disorders, DIF, PSI, Chi-square, Fit residual

INTRODUCTION

Quantifying pain and disability using patient-reported outcome measures (PROM) is an integral part in the evaluation of patients with any health condition. PROMs can be used to assess patient status, help set treatment goals and expectations; and more commonly to assess change following treatment interventions.[1] PROMs are used to assess outcomes in routine clinical practice, clinical research, and treatment trials because they provide a patient centered perspective which may differ from that provided by clinician based outcome measures (CBO).[2-5] Currently, there are two different approaches to assessment of clinical measurement properties of rating scales 1) Traditional psychometric methods[6] and 2) Modern methods (Rasch analysis[7] and Item response theory[8]). It has been suggested that Rasch analysis has a greater potential to identify the strengths and weaknesses of rating scales than traditional psychometric methods.[9]

A majority of currently available PROMs were developed prior to widespread use of Rasch and exist as an ordinal scale.[10] Issues have been raised with respect to the ability of these ordinal scales to provide a true quantitative scale that represents patient status along a continuum.[10-12] Forrest and Anderson (1986)[10] reported that when several items are measured on ordinal scales it is far from certain that the sum of scores has even ordinal properties. Merbitz et al. (1989)[12] commented that ordinal scales of measurement do not support the mathematical operations needed to calculate means and standard deviations. One of the most important assumptions of parametric analysis is that the variables must have been measured on the interval scale, so that it is possible to interpret the results.[13] The Rasch model provides a potential solution by providing a means to transform non-linear ordinal score to a more linear interval score, thus making the interpretation of the results possible and meaningful.

The Patient Rated Elbow Evaluation form (PREE)[14, 15] is a 20 item self-report measure, consisting of two sections, pain and function and the function section has two sub sections- ‘specific activities’ and ‘usual activities’. Responses are rated on a numeric rating scale. The pain section has five items of which four of them rate pain from ‘no pain’ (0) to ‘worst ever’ (10). The fifth item rates how often the patient has pain with responses ranging from ‘never’ (0) to ‘always’ (10). The responses on the function scale are anchored at ‘no difficulty’ (0) and ‘unable to do’ (10). The function section has 15 items regarding personal care, household

work, occupation and recreational activities out of which 11 items fall under the specific activities sub-section and 4 items are under the usual activities sub-section. All the scores are computed to obtain a global score out of 100. Higher PREE total scores reflect greater pain and disability.

The PREE has been considered a valid and reliable measure based on studies that have demonstrated strong clinical measurement properties using traditional psychometric methods. MacDermid et al (2001)[14] and Vincent et al (2013)[16] both reported construct validity with Pearson's correlations ranging from 0.49 to 0.84. The PREE has been found to have a very high level of internal consistency with Cronbach's alpha values above 0.90.[16, 17] In terms of longitudinal validity, the change scores correlated well with the change scores of other measures measuring similar construct.[16] In evaluating treatment effectiveness, the PREE has been reported to be highly responsive with large effect sizes and SRMs of 1.6 and 1.7 respectively.[16]

Angst et al (2005)[18] analyzed the factor structure of the PREE along with the patient component of the ASES-e and the DASH and have reported three principal components explaining around 89.2% of variance, of which the component 'physical specific' to which the PREE items loaded the maximum explains around 60.1% of variance. Vincent et al (2013)[16] reported an exploratory factor analysis, with principal component analysis. Four main components of the PREE were identified explaining about 77.2% of variance of the questionnaire's total score. These factors supported that the pain and usual items separated into individual subscales; items within the specific activities subscale separated into two components reflecting light and heavy activities.

Rasch analysis is a relatively recent addition to the family of analyses used to test the psychometric properties of rating scales. Rasch analysis is the formal testing of how well items and questionnaires follow axioms of clinical measurement that are linked to a mathematical measurement model called the Rasch model.[19] During Rasch analysis, responses from a set of individual questions from a questionnaire can be tested against response patterns predicted by the model. The pattern expected by the model is a deterministic pattern that follows a strict hierarchical ordering of items called Guttman scaling.[20] The PREE has been validated using traditional psychometric methods [14, 16-18] and has not been subjected to Rasch analysis which

means that interval level scaling has not been verified. Further, the potential for bias in different types of respondents has not been evaluated. Most studies using the PREE must assume interval level scaling or that parametric statistics are so robust that this will not affect results, since most rely on parametric statistics to make their conclusions, lack of interval level scaling or differential item functioning may lead to incorrect estimation of effects or false study conclusions.

Hence the purpose of the study was to conduct a Rasch analysis of the PREE to assess the overall fit to the Rasch model, the response scale used, individual item fit, differential item functioning (DIF), local dependency, unidimensionality and person separation.

METHODS

Research design: Cross sectional study using Rasch analysis

Participants

PREE scores were extracted from the charts of a cohort of 236 patients (Age range 21 - 79 years) who had completed outcome evaluations during surgical management of a variety of elbow conditions at the Roth MacFarlane Hand and Upper Limb Centre at St Joseph's Healthcare in London, Ontario. The cohort included patients who have undergone biceps tendon repair, total elbow replacement arthroplasty, radial head fixation and radial head arthroplasty. Subjects were included in the study if they underwent a surgical intervention for elbow pathology, were aged 20 and above and had completed the PREE. Subjects with cognitive impairment and communication difficulties due to neurologic or psychiatric disorders were excluded from the study.

Procedures

We selected the baseline post-operative data point to conduct a cross-sectional analysis since this time point is commonly used in assessment; and we anticipate that there would be substantial variability in patient responses.

RASCH ANALYSIS

Rasch analysis was performed using the RUMM 2030 software.[21] The 3 subscales of the PREE were analyzed separately for sources of misfit to the model using the analysis listed below. Since multiple testing was done, Bonferroni corrections were applied throughout the analyses as an adjustment. The steps laid out by Tenant and colleagues were followed. [22]

Likelihood ratio test: There are 2 types of Rasch models that can be used with a polytomous dataset. They are the rating scale model [23] and the partial credit model.[24] The rating scale model constrains all thresholds of responses to be equally spaced across the trait for all of the items.[25] The partial credit model places no constraints on the threshold parameters.[24] To determine which model to use we first performed a formal test called the Likelihood-Ratio Test.[26] If the result of this test is not significant then the rating scale model would be used and if the result is significant then the partial credit model will be used.[22] We used a partial credit model based on a significant likelihood ratio test.

Inspection of class interval structure: The number of class intervals and the distribution of persons were inspected by looking for intervals to be approximately equally distributed.[22]

Examination of the Thresholds: Category probability curves help us to examine responses to an item.[27-30] Examination of the category probability curves can reveal disordered thresholds, meaning inconsistent use of response items by the respondents. This is a common source of item misfit. Disordered thresholds occur when respondents have difficulty consistently discriminating between response options.[31] Potential solutions for correcting disordered thresholds include collapsing of the categories to improve the overall fit to the model.[32]

Fit statistics: The following important fit statistics are inspected when assessing the fit of the data to the Rasch model.

Item/person fit residuals: This tests the degree to which the Guttman pattern is achieved.[33] If the items and persons fit the model, we would observe a mean of approximately zero and a standard deviation of 1. The individual item and person–fit statistics are expressed as residuals. To say that the item and person fit the model we expect the residuals to range between $\pm$ 2.5.[26]

Item-trait interaction: This is tested to assess the property of invariance across the trait and is reported as a chi-square.[34] If the chi-square value is significant, this supports the presence of variance across the trait for hierarchical ordering of the items, compromising the required property of invariance.[26, 35]

Reliability indices: The Person-Separation-Index (PSI) [36-38] indicates the ability of the construct to discriminate amongst the respondents. The value of 0.7 is considered by convention to be the minimum acceptable level of PSI. The PSI determines the number of groups of patients between whom we can statistically differentiate. A value of 0.8 is representing the ability to statistically differentiate at least 3 groups. A value of 0.9 would indicate the ability to discriminate between 4 or more groups.[38, 39] PSI is an indicator of how much we can rely on the fit characteristics.[38] Lower PSI indicates less reliability.[37]

Differential item functioning (DIF): DIF is another potential source of item bias resulting in misfit of the data to the model. Despite having the balance of underlying characteristics, different groups can respond in a different way to the same item. DIF can be detected graphically (Item characteristic curves) and statistically (ANOVA). Uniform DIF is indicated by a significant main effect for the person factor (gender in this case), while the presence of non-uniform DIF is indicated by a significant interaction effect (gender x class interval).[32] There are 2 types of DIF- a) Uniform DIF, where the group shows a consistent systematic difference in their responses to an item, across the whole range of the attribute being measured; b) When there is non-uniformity in the differences between the groups (e.g. it varies across levels of the attribute) then this is referred to as non-uniform DIF.[32] With Uniform DIF the problem can be remedied by splitting the file by group and separately calibrating the item for each group. Non-uniform DIF is more problematic because there is no mathematical adjustment; and typically it would require removing the item from the scale.[22] We assessed DIF for gender and age groups.

Local dependency: A violation of local independence occurs when examined item responses depend not just on their trait level, but on their responses to other test items.[40] Principal component analysis (PCP) [41, 42] of the residuals was done as a test for local independence. An inter-item residual correlation $\geq$ 0.3 was used as a cut-off to indicate local dependency. [40] The residuals were inspected visually and the lack of any meaningful pattern in was taken as an indicator of local independence and consequently unidimensionality of the scale. [43]

Unidimensionality: This was formally tested by the method proposed by Smith where we allow the factor loadings on the first residual to determine subsets of items and then testing, by a paired t test, to see if the person estimate derived from these subsets significantly differs from that derived from all items.[43] We expect the percentage of tests that are significant ($P<0.05$) should be less than 5%, for the questionnaire to be unidimensional.

Targeting: Every questionnaire should be well-targeted towards the patient population in question. In other words the difficulty of the items of a questionnaire should be slightly higher than the ability of the targeted population. This was analyzed by plotting the person-item location threshold distribution graph with distributions of persons on the top half of the graph and item thresholds at the bottom half of the graph. The average person location should be zero logits.[22]

RESULTS

There were no missing data and all 236 cases were determined to be valid by the RUMM 2030 software. The 3 sub scales were analysed separately. The class intervals were checked throughout the analysis for consistency and the cases were nearly equally distributed between the groups. (See table 1, initial analysis)

Handling of data to fit the Rasch model:

Pain subscale: Analysis of the 5 items of the pain sub scale revealed slight deviation from the Rasch model requirements as indicated by a high and significant item trait interaction ($p<0.001$). (See table 1) Items 3, 4 and 5 exhibited disordered thresholds. Individual item fit was excellent indicating acceptable levels of discrimination.(See table 3) Uniform DIF for age group was observed for item 2, "*Pain* - At rest". (See Table 4) Unidimensionality was acceptable. (See Table 1; initial analysis) The Reliability Index was high with a PSI of 0.87. No meaningful pattern of local dependency was observed.

To improve the overall fit to the Rasch model items 3, 4, and 5 were rescored. Then item 2 was split for age group; this resulted in excellent item fit and non-significant item trait interaction. Uniform DIF (Age group) for item 2 was not evident. (Table 5) Unidimensionality was observed and no local dependency was present. The reliability improved to be 0.90. (See Table 1; final analysis) Targeting was also good as indicated by the person item threshold map.

(See Figure 2 A) This implies that this sub scale has a good coverage for elbow disorders related to pain. Hence, this was accepted as the final model.

Specific activities subscale: Rasch analysis revealed that the 11 item specific activities subscale has marked deviations from the Rasch model expectations. This was evident from, the disordered thresholds (11 out of the 11 items); the property of invariance was compromised because of large and significant chi square value that was observed. There was a breach of unidimensionality as well. (See Table 1; initial analysis) Local dependency was observed between the following items, Item 1 "*Comb my hair*"; Item 2 "*Eat with a fork or spoon*"; Item 3 "Pull a heavy object"; Item 5 "*Carry a 10lb object with my arm at my side*"; Item 9 "*Wash my opposite armpit*"; Item 10 "*Tie my shoe*". The only meaningful pattern observed was for item 3 and 5 that were about heavy objects. DIF analysis revealed that none of the items exhibited DIF for age group or gender. (See table 4). Individual item fit was excellent indicating acceptable levels of discrimination.(See table 3) and the reliability of the scale was high (PSI = 0.83). (See Table 1; initial analysis)

To improve the fit of the specific activities subscale to Rasch model various actions were taken. Initially the 11 items with disordered thresholds were rescored. Thresholds were more disordered in the middle of the 0-10 scale. So categories were collapsed to a 5 point or a 6 point scale depending on the item. (See Table 2; Figure 1) To deal with local dependency, subtest analyses was done to see if they cancelled at the sub test level. Testlets were created by combining items 1 and 2; 3, 4 and 5; and 8, 9 and 10. When the subtest analysis was completed local dependency was cancelled, the chi square residual became non-significant indicating acceptable fit of the data to the Rasch model. Unidimensionality was observed. The reliability improved to be 0.91. (See Table 1; final analysis) Targeting was acceptable with enough coverage (see figure 2 B)

Usual activities subscale: The usual activities subscale initially demonstrated misfit to the Rasch model with disordered thresholds for three of the four items (items1, 3 and 4). There was no DIF for age group. Uniform DIF for gender was observed for item 2 "*Household work (cleaning, maintenance)"*.(see Table 4) There was no breach of the properties of invariance, local independence and unidimensionality. Reliability was acceptable (PSI = 0.82). (See table 1; initial analysis)

To improve the fit of the scale to the Rasch model the items with disordered thresholds were rescored to reorder them. (See table 2). To deal with DIF for gender, item 2 was split for gender. The final analysis rendered the data to fit the Rasch model, increasing reliability of the sub scale (PSI = 0.86) and bringing down the chi square value. (See table 1; final analysis) The scale was well targeted as demonstrated by the person-item threshold map. (See figure 2 C)

DISCUSSION

The results of this Rasch analysis support the claims made by classical test methods on the psychometric properties of the PREE. [16] It has pointed to areas of improvement in scoring for the PREE to derive an unbiased patient reported estimate of pain and disability in elbow disorders.

The Rasch measurement model provides a relatively recent approach to development and evaluation of clinical measures and provides an in depth analysis of measurement traits on aspects less attended by previous psychometric analyses. Since many currently used measures pre-date the common use of Rasch- they may not necessarily have been developed in a way that fulfills the requirements of Rasch model. The PREE however, exhibited acceptable level of fit to the Rasch model requirements with less complicated data handling. By assessing the fit of the PREE data to the Rasch model, and following a sequential Rasch approach to assess potential sources of misfit we have identified areas that need to be improved to achieve a linear interval score. These interval scores can accurately reflect change in patient disability status; whereas an ordinal scale cannot.[13]

The PREE had 17 items (3 items from the pain sub scale; 11 from the specific activities; 3 from usual activities) with disordered thresholds out of the 20 items. This draws our attention to the 0 to 10 numeric rating scale (an ordinal scale) used in this self-report measure. Similar findings have been observed in the Patient Rated Wrist and Hand Evaluation (PRWHE), the wrist and hand counterpart of the PREE.[44] One of the possible reasons for this large number of disordered thresholds could be that there are too many options for the patients to choose from, that is, there are too many options for patients to calibrate. [23, 45] Another possibility is that the questions are too difficult for the patients to answer. However, during development of this measure patients preferred the 0-10 scale as they it found easier; and found the items easy to

understand. [14] Furthermore, the PREE was shown to be well-targeted with a person-item location slightly less than the average of zero logits, which discounts item difficulty as a problem. (See figure 2) Rescoring of these items as indicated in table 2 places additional burden on the clinician but may retain ease of administration. Alternatively the scaling can be redesigned to be a 6 point (0-5) scale. However, it might be challenging to select the right descriptors for this scale. Further, the 0-10 scale is commonly used in clinical practice and is more sensitive and easily understood than VAS scales. [46] Finally, such a substantial change on the basis of one study might be preliminary- particularly since changes to scoring were able to address most measurement concerns. Therefore, it seems that rather than changing the scale, changing the scoring is a preferable approach.

In all the three sub scales none of the items demonstrated a misfit as indicated by fit residuals that were within acceptable limits. (See Table 3) This indicates that none of the items were over discriminating. We observed a large and significant chi square initially for the specific activities sub scale indicating the presence of a latent trait violating the property of invariance. However this got adjusted when testlets were created in the sub test analysis. Our PSI was acceptable (PSI> 0.80) indicating that our sample size was adequate for this Rasch analysis.

To satisfy the assumptions of unidimensionality it is suggested that three sub scales of the PREE be considered separately. Scoring pain and disability subscales separately aligned with the developers original intention of having these subscales and establishing scale reliability,[14] and is in agreement with recommendations for the similar PRWE both based on Rasch analysis[44] and expert consensus.[47] However, many studies continue to report the total score of pain and disability measures, perhaps because having a single primary outcome measure is preferred for study design and interpretation. Where such a composite score is used, the user should be careful to analyse the deconstructed measure and insure that conclusions are not affected by pooling.

Unidimensionality was not an issue with the pain and usual activities subscale. However, the specific activities subscale exhibited multidimensionality. This confirms the observations made through an exploratory factor analysis where the specific activities subscale loaded onto more than 1 factor.[16] In the current study the cause for multidimensionality was local dependency observed between the items. However this local dependency cancelled out when subtest analysis was performed. This indicates that there are some redundant items in the specific

activities subscale that could potentially be reduced. Since the measure is established and brief, the benefits of this would need consideration.

In the pain sub scale of the PREE the item “Pain: At rest” was the source of misfit. This item demonstrated a uniform DIF for age group. This is not surprising as previous basic science research findings indicate that pain tolerance is reduced as people age and this means there is a possibility of older people perceiving their pain levels different from the younger ones, which is evident in our sample.[48-50] We identify this as a potential area of concern that should be tested in other samples and with larger sample sizes.

Uniform DIF for gender was observed for the item “Household work (cleaning, maintenance)” ($p=0.001$) under usual activities of the function subscale. There can be gender-based differences in “household work" with men usually performing heavier household work while women tend to do lighter work [51] or a greater portion of the work. [52] This may explain why men and women answered this question differently. Gender was considered as a potential source of differential response when designing this scale (which pre-dated Rasch). [14] We recommend future studies to look into the gender differences in responding to the PREE items. Since we only examined differential item functioning based on gender and age group, there is a need to conduct examination for other potential sources including affected side. More clinical constructs can be added to the DIF analysis to see how the individual items are behaving with the different constructs. Since gender and age are commonly reported in clinical research studies, the distributions of these may need to be considered when interpreting the PROM reported in clinical studies in patients with elbow conditions that use the PREE or other measures where Rasch has not been used to insure interval level scaling.

With the increasing use of Rasch, new flaws are being detected in many PROMs that were developed using more traditional clinimetric approaches. This has potential to improve clinical measurement by improving or discarding tools that do not provide valid measurement. However, we suggest a cautious approach in suggesting changes to measures. Different Rasch analyses on the same scale across different studies have reported different findings and made different recommendations about what changes should be made. We found that changing the scale scoring to meet Rasch based interval level scaling can have an impact on study conclusions, [53] but few others have undertaken such evaluations when proposing that scores

need to be changed. When the threshold for changing PROM is low, this can result in a range of published PROM variants with no clear choice of the best option. The potential benefits to change the scale must be weighed against the well documented knowledge translation challenges in implementation of PROM [54] and need for consistency across comparisons. Hence, we suggest that where findings are consistent with previous psychometric findings and support the current PREE (with item rescoring) then this warrants continued use of the current PREE. Where we have found suboptimal measurement findings that are not consistent with that reported in other studies or across time-points we suggest caution and further study.

The strengths of the current study are its high PSI values, high Cronbach's alpha and the excellent power of fit with a sample size of 236 patients. The limitations of the current study are: not all elbow disorders were represented and that we looked at DIF only for gender and age. We recommend future studies include a variety of elbow disorder patients; evaluate other potential sources of differential item functioning such as occupational demand, severity of injury, level of education, Worker's Compensation Insurance Board (WCIB) claims and other social factors that might determine the DIF. Our findings questioned the measurement properties of the items of the specific activities subscale. It might be worthwhile exploring the stability of our findings before implementing substantial changes- particularly in light of the strong psychometric properties demonstrated in previous studies using classical test methods. In one of our previous studies we found that the PREE has demonstrated excellent internal consistency, construct validity, longitudinal validity and sensitivity to change.[16] Angst et al have also found the PREE to demonstrate acceptable levels of validity, reliability and responsiveness. [18]

CONCLUSION

All the three sub scales of the PREE appear to be robust when tested against the Rasch model amenable to few changes. Rasch analysis has highlighted areas needing further investigations and potential modification of the rating scale due to the misfit caused by disordered thresholds in our sample. Additional studies are needed to assess the optimal format and scoring of the PREE.

Competing Interests: The authors declare that none of them have any competing interests.

Author contributions: JV conceived of the study, responsible for study design, execution of the study, performed the statistical analysis and drafted the initial and completed all revision of the

manuscript. JMD participated in the design of the study, coordinated data collection, helped with interpretation of the statistical analysis and helped to draft/revise the manuscript. RG and GJK enrolled subjects, contributed to study design/interpretation and revised the drafts of the manuscript. All authors read and approved the final manuscript.

REFERENCES

1. MacDermid JC, Stratford P: **Applying Evidence on Outcome Measures to Hand Therapy Practice.** Journal of Hand Therapy 2004, **17**(2):165-73.

2. Food and Drug Administration: **Patient reported outcome measures: use in medical product development to support labeling claims.** 2009, **2011**(11/8/2011).

3. Lieberman JR, Dorey F, Shekelle P, Schumacher L, Thomas BJ, Kilgus DJ, Finerman GA: **Differences between patients' and physicians' evaluations of outcome after total hip arthroplasty.** J Bone Joint Surg Am 1996, **78**(6):835-838.

4. Slevin ML, Plant H, Lynch D, Drinkwater J, Gregory WM: **Who should measure quality of life, the doctor or the patient?** Br J Cancer 1988, **57**(1):109-112.

5. Sprangers MA, Aaronson NK: **The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: a review.** J Clin Epidemiol 1992, **45**(7):743-760.

6. Nunnally JC, Bernstein IH: **Psychometric theory***:* 3rd ed. New York: McGraw-Hill; 1994.

7. Andrich D: **Rasch models for measurement***:* Newbury Park: Sage Publications; 1988.

8. DeMars C: **Item response theory***:* Oxford ; New York: Oxford University Press; 2010.

9. Hobart J, Cano S: **Improving the evaluation of therapeutic intervention in MS: the role of new psychometric methods.**

10 UK Health Technology Assessment Programme V*ol 13.* 2009

10. Forrest M, Andersen B: **Ordinal scale and statistics in medical research.** Br Med J (Clin Res Ed) 1986, **292**(6519):537-538.

11. Horton M, Tennant A: **Patient Reported Outcomes: misinference from ordinal scales?** Trials 2011, **12**(Suppl 1):A65.

12. Merbitz C, Morris J, Grip JC: **Ordinal scales and foundations of misinference.** Arch Phys Med Rehabil 1989, **70**(4):308-312.

13. Siegel S, Castellan NJ: **Nonparametric statistics for the behavioral sciences***:* 2nd ed. New York ;; Montreal: McGraw-Hill; 1988.

14. MacDermid JC: **Outcome evaluation in patients with elbow pathology: issues in instrument development and evaluation.** J Hand Ther 2001, **14**(2):105-114.

15. MacDermid JC: **Patient Rated Elbow Evaluation.** [http://www.srs-mcmaster.ca/Portals/20/pdf/research_resources/PREE.pdf.]

16. Vincent JI, Macdermid JC, King GJ, Grewal R: **Validity and sensitivity to change of patient-reported pain and disability measures for elbow pathologies.** J Orthop Sports Phys Ther 2013, **43**(4):263-274.

17. John M, Angst F, Pap G, Junge A, Mannion AF: **Cross-cultural adaptation, reliability and validity of the Patient Rated Elbow Evaluation (PREE) for German-speaking patients.** Clin Exp Rheumatol 2007, **25**(2):195-205.

18. Angst F, John M, Pap G, Mannion AF, Herren DB, Flury M, Aeschlimann A, Schwyzer HK, Simmen BR: **Comprehensive assessment of clinical outcome and quality of life after total elbow arthroplasty.** Arthritis Rheum 2005, **53**(1):73-82.

19. Rasch G: **Probabilistic models for some intelligence and attainment tests***:* Chicago: University of Chicago Press; 1960.

20. Akkermans W: **Polytomous item scores and Guttman dependence.** British Journal of Mathematical & Statistical Psychology 1999, **52**(00071102):39-61.

21. Andrich D, Sheridan B, Luo G: **RUMM 2030.** *Perth, WA: RUMM Laboratory Pty Ltd* 2010,

22. Tennant A, Horton MC, Pallant JF: **Introductory Rasch Analysis: A workbook.** Department of Rehabilitation Medicine, University of Leeds, UK, 2011, .

23. Andrich D: **A rating formulation for ordered response categories.** Psychometrika 1978, **43**(4):561-573.

24. Masters G: **A rasch model for partial credit scoring.** Psychometrika 1982, **47**(2):149-174.

25. Wright BD, Masters GN: **Rating scale analysis***:* Chicago: MESA Press; 1982.

26. Wright BD, Stone MH: **Best Test Design. Rasch Measurement.** In *The measurement model.* Edited by Anonymous Chicago: Mesa press; 1979:1-17.

27. Andrich D: **Category ordering and their utility.** Rasch Measurement Transactions 1996, **9**(4):464-465.

28. Linacre JM: **Step disordering and Thurstone thresholds.** Rasch Measurement Transactions 1991, **5**(3):171.

29. Linacre JM: **Investigating rating scale category utility.** J Outcome Meas 1999, **3**(2):103-122.

30. Linacre JM: **Optimizing rating scale category effectiveness.** J Appl Meas 2002, **3**(1):85-106.

31. Zhu W, Updyke WF, Lewandowski C: **Post-hoc Rasch analysis of optimal categorization of an ordered-response scale.** J Outcome Meas 1997, **1**(4):286-304.

32. Pallant JF, Tennant A: **An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS).** Br J Clin Psychol 2007, **46**(Pt 1):1-18.

33. Smith RM: **IPARM: Item and Person analysis with the Rasch model***:* Chicago: MESA Press; 1991.

34. Linacre,J.M., Wright, B. D.: **(Dichotomous mean square) Infit and Outfit Chi-square fit statistics.** Rasch Measurement Transactions 1994, **8**(2):360-361.

35. Romanoski J, Douglas G: **Test scores, measurement, and the use of analysis of variance: an historical overview.** J Appl Meas 2002, **3**(3):232-242.

36. Andrich D: **An index of person separation in latent trait theory, the traditional KR20 index, and the Guttman scale response pattern.** Educ Psychol Res 1982, **9**:95-104.

37. Wright BD: ***Reliability and separation*.** Rasch Measurement Transactions 1996, **9**(4):472.

38. Wright BD: **Interpreting reliabilities.** Rasch Measurement Transactions 1998, **11**(4):602.

39. Fisher W: **Reliability statistics.** Rasch Measurement Transactions 1992, **6**(3):238.

40. Embretson SE, Reise SP: **Item response theory for psychologists***:* Mahwah, N.J.: Lawrence Erlbaum Associates, Publishers; 2000.

41. Linacre JM: **Structure in Rasch residuals: Why principal component analysis?** Rasch Measurement Transactions 1998, **12**(2):636.

42. Linacre JM: **Detecting multidimensionality: which residual data-type works best?** J Outcome Meas 1998, **2**(3):266-283.

43. Smith EV,Jr: **Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals.** J Appl Meas 2002, **3**(2):205-231.

44. Packham T, MacDermid JC: **Measurement properties of the Patient-Rated Wrist and Hand Evaluation: Rasch analysis of responses from a traumatic hand injury population.** J Hand Ther 2013, **26**(3):216-23; quiz 224.

45. Preston CC, Colman AM: **Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences.** Acta Psychol (Amst) 2000, **104**(1):1-15.

46. Williamson A, Hoggart B: **Pain: a review of three commonly used pain rating scales.** J Clin Nurs 2005, **14**(7):798-804.

47. Goldhahn J, Beaton D, Ladd A, Macdermid J, Hoang-Kim A: **Recommendation for measuring clinical outcome in distal radius fractures: a core set of domains for standardized reporting in clinical practice and research.** Arch Orthop Trauma Surg 2014, **134**(2):197-205.

48. Lautenbacher S, Kunz M, Strate P, Nielsen J, Arendt-Nielsen L: **Age effects on pain thresholds, temporal summation and spatial summation of heat and pressure pain.** Pain 2005, **115**(3):410-418.

49. Gibson SJ, Helme RD: **Age-related differences in pain perception and report.** Clin Geriatr Med 2001, **17**(3):433-56, v-vi.

50. Cole LJ, Farrell MJ, Gibson SJ, Egan GF: **Age-related differences in pain sensitivity and regional brain activity evoked by noxious pressure.** Neurobiol Aging 2010, **31**(3):494-503.

51. Bianchi SM, Milkie MA, Sayer LC, Robinson JP: **Is Anyone Doing the Housework? Trends in the Gender Division of Household Labor.** Social Forces 2000, **79**(1):191-228.

52. Miller A, Sassler S: **The Construction of Gender Among Working-Class Cohabiting Couples.** Qualitative Sociology 2012, **35**(4):427-446.

53. Cheifetz O, Packham TL, Macdermid JC: **Rasch analysis of the Edmonton Symptom Assessment System and research implications.** Curr Oncol 2014, **21**(2):e186-94.

54. Snyder C, Aaronson N, Choucair A, Elliott T, Greenhalgh J, Halyard M, Hess R, Miller D, Reeve B, Santana M: **Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations.** Quality of Life Research 2012, **21**(8):1305-1314.

Table 1: Summary fit statistics for individual subscales of the PREE

Analysis	Item fit residual		Person Fit residual		Item-trait interaction		Unidimensionality	**PSI**
	Mean	SD	Mean	SD	Chi square *(df)*	*P*	Per C<5% (95% C.I)	
PAIN SUBSCALE								
Initial	-0.01	1.02	-0.34	0.88	13.77 (15)	0.54	0.05 (0.02 – 0.07)	0.90
Final	-0.09	0.94	-0.34	0.87	17.87 (18)	0.47	- (since items were split for DIF)	0.90
SPECIFIC ACTIVITIES SUBSCALE								
Initial	0.08	1.27	-0.32	1.21	55.96 (33)	0.01*	0.1 (0.07 – 0.13)*	0.90
Final	0.05	1.82	-0.39	1.08	10.87 (15)	0.76	0.03 (0.01 – 0.06)	0.91
USUAL ACTIVITIES SUBSCALE								
Initial	-0.72	0.56	-0.41	0.94	13.39 (12)	0.34	0.02 (0.01 – 0.05)	0.82
Final	-0.55	1.02	-0.44	1.02	6.01 (10)	0.82	- (since items were split for DIF)	0.86

*Source of misfit to the Rasch model; SD = Standard deviation; df = Degrees of freedom; per C<5% = proportion of t tests that were significant at level of significance of 0.05; 95% CI = 95% confidence interval; PSI = Person separation index; PREE – Patient Rated elbow Evaluation
For the data to satisfy Rasch model requirements:

- Mean is expected to be approx. around zero (Can range between $\pm$ 2.5);
- S.D. should be approx. 1;
- Chi square value is expected to be small and statistically non-significant;
- For a measure to be unidimensional per C<5% should be less than 0.05; if it is higher than 0.05 then look into the lower limit the 95% confidence interval if it is less than 0.05 then the measure is unidimensional.
- PSI (Person separation index) should be greater than 0.70 for the summary statistics to be reliable;

Table 2: Table showing the structure of scores for individual items of the PREE

Item	0	1	2	3	4	5	6	7	8	9	10
PAIN SUB SCALE											
Pain - When it is at its worst	0	1	2	3	4	5	6	7	8	9	10
Pain - At rest	0	1	2	3	4	5	6	7	8	9	10
Pain - When lifting a heavy object*	0	1	1	2	3	4	5	5	6	6	7
Pain - When doing a task with repeated elbow movement*	0	1	2	3	4	4	5	5	6	6	7
How often do you have pain?*	0	1	2	3	4	4	5	5	6	6	7
SPECIFIC ACTIVITIES SUB SCALE											
Comb my hair*	0	1	1	1	2	2	2	3	3	3	4
Eat with a fork or spoon*	0	1	1	1	2	2	3	3	3	3	4
Pull a heavy object*	0	1	1	1	2	2	2	3	3	3	4
Use my arm to rise from a chair*	0	1	1	2	2	3	3	4	4	4	5
Carry a 10lb object with my arm at my side*	0	1	2	2	3	3	4	4	4	4	5
Throw a small object, such as a tennis ball*	0	1	1	2	2	2	3	3	3	3	4
Use a telephone*	0	1	1	2	2	3	3	3	4	4	5
Do up buttons on the front of my shirt*	0	1	1	2	2	3	3	3	4	4	5
Wash my opposite armpit*	0	1	1	1	2	2	2	3	3	3	4
Tie my shoe*	0	1	1	1	1	2	2	2	2	2	3
Turn the doorknob and open a door*	0	1	1	2	2	2	2	3	3	3	4
USUAL ACTIVITIES SUB SCALE											
Personal activities (dressing, washing)*	0	1	1	2	2	3	3	4	5	6	7
Household work (cleaning, maintenance)	0	1	2	3	4	5	6	7	8	9	10
Work (your job or everyday work)*	0	1	1	2	2	3	3	4	4	4	5
Recreational activities*	0	1	1	2	2	3	3	3	4	4	5

*Rescored items; PREE – Patient Rated elbow Evaluation

Table 3: Initial fit statistics for individual items of the PREE

Item	Location	SE	Fit statistics		
			Fit residual	Chi Square	Chi square probability
PAIN SUB SCALE					
Pain - When it is at its worst	-0.32	0.05	0.88	1.62	0.66
Pain - At rest	1.32	0.04	-1.10	2.05	0.56
Pain - When lifting a heavy object	-0.67	0.06	1.05	3.85	0.28
Pain - When doing a task with repeated elbow movement	-0.36	0.05	-1.28	4.35	0.23
How often do you have pain?	0.03	0.06	0.06	1.87	0.60
SPECIFIC ACTIVITIES SUB SCALE					
Comb my hair	0.50	0.08	-1.42	7.70	0.05
Eat with a fork or spoon	0.55	0.08	-1.26	4.15	0.25
Pull a heavy object	-1.25	0.09	-0.91	1.07	0.78
Use my arm to rise from a chair	-0.35	0.07	-0.12	3.08	0.38
Carry a 10lb object with my arm at my side	-0.82	0.07	2.29	13.61	0.003
Throw a small object, such as a tennis ball	-0.60	0.08	0.96	0.24	0.97
Use a telephone	0.63	0.07	1.03	5.69	0.13
Do up buttons on the front of my shirt	0.42	0.07	0.91	2.64	0.45
Wash my opposite armpit	0.44	0.08	-1.04	6.41	0.09
Tie my shoe	0.37	0.10	1.27	6.19	0.10
Turn the doorknob and open a door	0.10	0.08	0.56	1.43	0.70
USUAL ACTIVITIES SUB SCALE					
Personal activities (dressing, washing)	0.86	0.06	-0.97	1.91	0.39
Household work (cleaning, maintenance)	-0.04	0.05	-2.17	3.95	0.14
Work (your job or everyday work)	-0.35	0.07	0.25	2.01	0.37
Recreational activities	-0.47	0.07	0.65	0.12	0.94

SE-standard error; PREE – Patient Rated elbow Evaluation

Table 4: DIF summary (gender) for the individual items of the PREE

Item	Uniform DIF for Gender				Non-Uniform DIF for Gender			
	MS	F	DF	*P*	MS	F	DF	*P*
PAIN SUB SCALE								
Pain - When it is at its worst	3.76	4.31	1	0.04	0.88	1.01	3	0.39
Pain - At rest	0.83	1.28	1	0.26	1.00	1.53	3	0.21
Pain - When lifting a heavy object	0.00	0.00	1	0.96	0.10	0.10	3	0.96
Pain - When doing a task with repeated elbow movement	1.58	2.43	1	0.12	0.07	0.11	3	0.95
How often do you have pain?	0.40	0.50	1	0.48	0.50	0.62	3	0.61
SPECIFIC ACTIVITIES SUB SCALE								
Comb my hair	3.25	4.76	1	0.03	0.29	0.42	3	0.74
Eat with a fork or spoon	0.10	0.14	1	0.71	1.37	1.97	3	0.12
Pull a heavy object	0.00	0.00	1	0.97	0.15	0.19	3	0.90
Use my arm to rise from a chair	1.52	1.77	1	0.19	1.20	1.39	3	0.25
Carry a 10lb object with my arm at my side	0.11	0.08	1	0.78	0.79	0.60	3	0.61
Throw a small object, such as a tennis ball	3.34	3.35	1	0.07	3.08	3.08	3	0.03
Use a telephone	0.90	0.84	1	0.36	0.53	0.49	3	0.69
Do up buttons on the front of my shirt	0.04	0.04	1	0.84	2.73	2.72	3	0.05
Wash my opposite armpit	0.51	0.67	1	0.41	0.37	0.49	3	0.69
Tie my shoe	0.49	0.47	1	0.49	0.20	0.19	3	0.91
Turn the doorknob and open a door	0.00	0.00	1	0.99	2.11	2.21	3	0.09
USUAL ACTIVITIES SUB SCALE								
Personal activities (dressing, washing)	0.65	1.03	1	0.31	-0.16	-0.25	2	0.99
Household work (cleaning, maintenance) *	6.68	14.28	1	0.00	-0.06	-0.14	2	0.99
Work (your job or everyday work)	3.24	4.35	1	0.04	0.72	0.97	2	0.38
Recreational activities	3.74	4.61	1	0.03	0.27	0.33	2	0.72

* Items exhibiting Uniform DIF. An item was considered to exhibit DIF if *P* values are significant after applying Bonferroni correction factor; PREE – Patient Rated Elbow Evaluation

Table 5: DIF summary (Age Group) for the individual items of the PREE

Item	Uniform DIF for Age				Non-Uniform DIF for Age			
	MS	F	DF	*P*	MS	F	DF	*P*
PAIN SUB SCALE								
Pain - When it is at its worst	1.55	1.77	3	0.15	0.92	1.05	9	0.40
Pain - At rest*	**2.98**	**4.82**	**3**	**0.00**	0.79	1.28	9	0.25
Pain - When lifting a heavy object	1.29	1.36	3	0.26	0.40	0.42	9	0.93
Pain - When doing a task with repeated elbow movement	1.31	2.02	3	0.11	0.44	0.68	9	0.73
How often do you have pain?	0.89	1.11	3	0.35	0.63	0.78	9	0.64
SPECIFIC ACTIVITIES SUB SCALE								
Comb my hair	0.38	0.58	3	0.63	1.38	2.08	9	0.03
Eat with a fork or spoon	0.62	0.90	3	0.50	0.84	1.20	9	0.30
Pull a heavy object	1.23	1.58	3	0.20	0.50	0.64	9	0.76
Use my arm to rise from a chair	1.25	1.41	3	0.24	0.45	0.51	9	0.86
Carry a 10lb object with my arm at my side	3.96	3.32	3	0.02	2.20	1.84	9	0.06
Throw a small object, such as a tennis ball	1.74	1.69	3	0.17	1.14	1.11	9	0.35
Use a telephone	0.99	0.94	3	0.42	0.94	0.89	9	0.54
Do up buttons on the front of my shirt	2.05	1.99	3	0.12	0.65	0.63	9	0.77
Wash my opposite armpit	1.94	2.54	3	0.06	0.21	0.27	9	0.98
Tie my shoe	0.29	0.27	3	0.84	0.57	0.54	9	0.85
Turn the doorknob and open a door	1.39	1.46	3	0.22	1.05	1.10	9	0.36
USUAL ACTIVITIES SUB SCALE								
Personal activities (dressing, washing)	0.07	0.10	3	0.96	0.12	0.19	6	0.98
Household work (cleaning, maintenance)	1.38	2.83	3	0.04	0.33	0.67	6	0.67
Work (your job or everyday work)	0.95	1.28	3	0.28	1.33	1.81	6	0.10
Recreational activities	2.41	2.99	3	0.03	0.47	0.58	6	0.74

* Items exhibiting Uniform DIF. An item was considered to exhibit DIF if *P* values are significant after applying Bonferroni correction factor; PREE – Patient Rated Elbow Evaluation

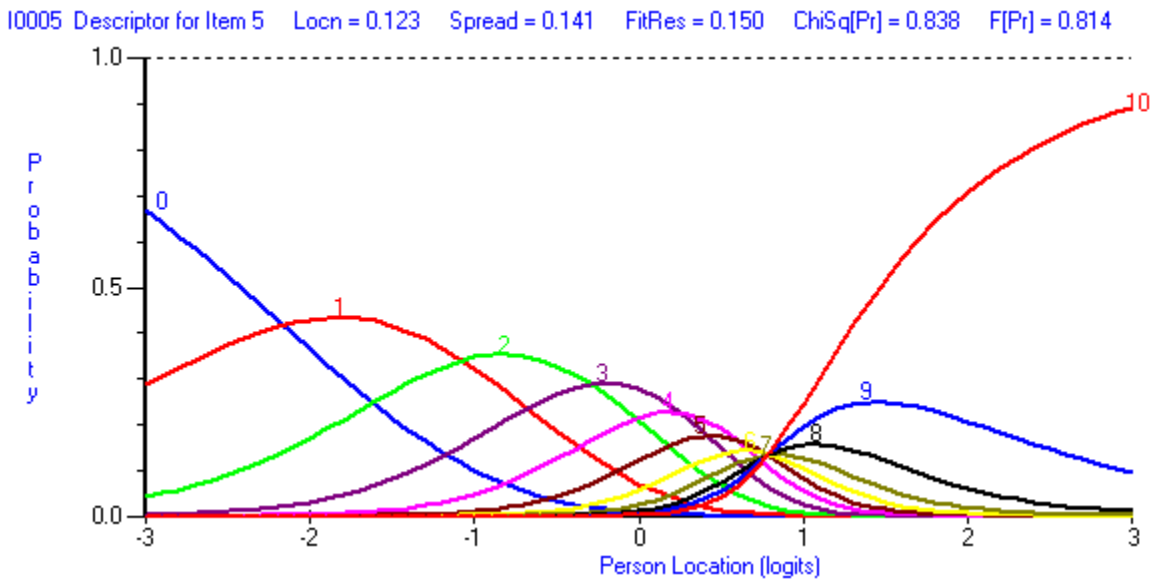
Table 6: Principal component analysis (PCP) showing first component loadings for individual items of the PREE

ITEM	PRINCIPAL COMPONENT 1
PAIN SUB SCALE	
Pain - When it is at its worst*	0.01
Pain - At rest	-0.69
Pain - When lifting a heavy object*	0.74
Pain - When doing a task with repeated elbow movement*	0.55
How often do you have pain?*	-0.43
SPECIFIC ACTIVITIES SUB SCALE	
Comb my hair*	0.12
Eat with a fork or spoon*	0.09
Pull a heavy object	-0.69
Use my arm to rise from a chair	-0.42
Carry a 10lb object with my arm at my side	-0.63
Throw a small object, such as a tennis ball	-0.33
Use a telephone*	0.35
Do up buttons on the front of my shirt*	0.74
Wash my opposite armpit*	0.59
Tie my shoe*	0.45
Turn the doorknob and open a door	-0.13
USUAL ACTIVITIES SUB SCALE	
Personal activities (dressing, washing)	-0.61
Household work (cleaning, maintenance)	-0.65
Work (your job or everyday work)*	0.68
Recreational activities*	0.72

*Positively loaded items; PREE – Patient Rated elbow Evaluation

Figure 1: Showing disordered threshold for item 5 "How often do you have pain?" of the pain subscale

A) Before rescoring



B) After rescoring

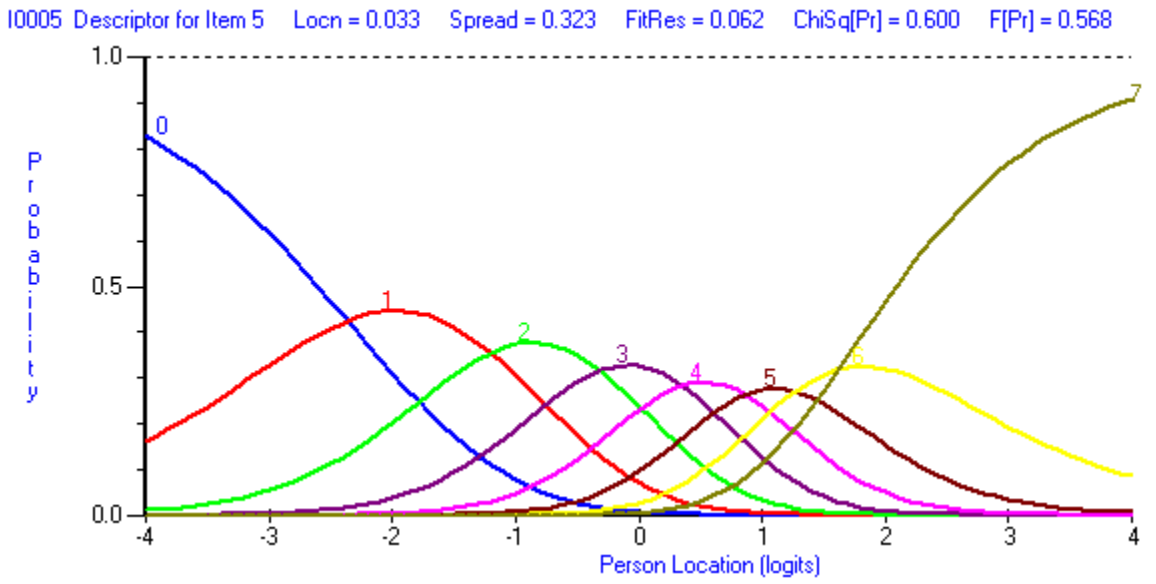
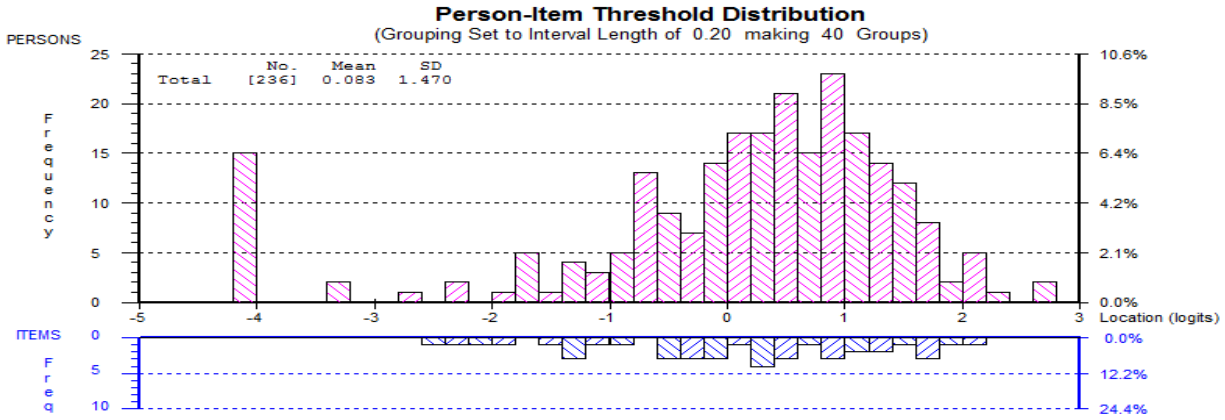
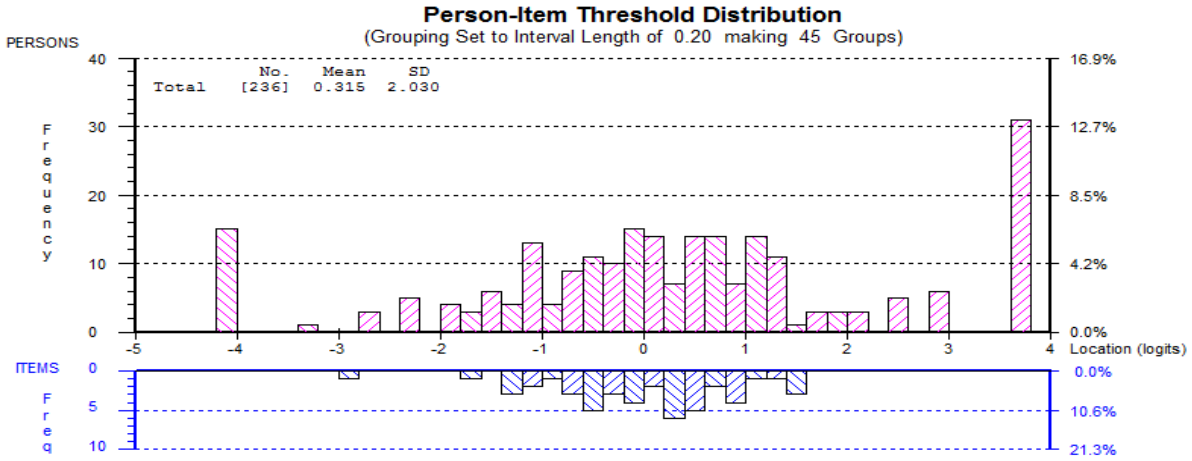


Figure 2: Person-item threshold distributions for the individual subscales of the Patient Rated Elbow Evaluation questionnaire showing targeting

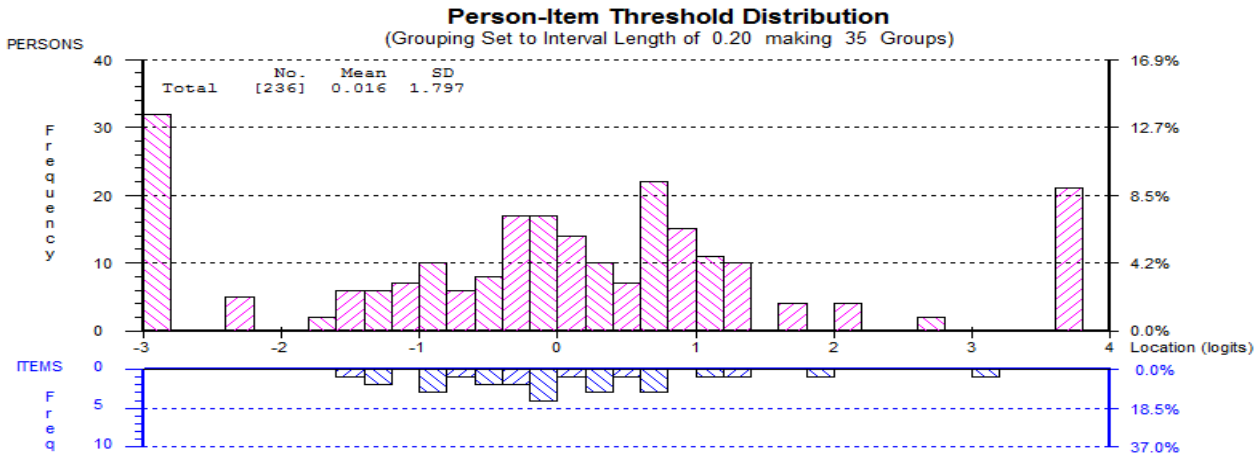
A) Pain Sub scale



B) Specific Activities sub scale



C) Usual Activities sub scale



CHAPTER 7: GENERAL DISCUSSION

Summary

In order to use a patient reported outcome measure (PROM) in routine clinical practice it is important to understand the reliability, validity, responsiveness, cost and utility of the measure. The PREE and the pASES-e have been recommended as candidates for assessment of pain and disability alongside other measures addressing different aspects of assessment of the elbow, when a comprehensive assessment of the elbow is performed by a clinician.[1,2] We conducted a thorough literature review on the clinical measurement properties of these two measures and identified potential gaps in the literature. Our purpose was to evaluate the psychometric properties of the PREE and the pASES-e and to fill some of the current gaps in the literature, while acknowledging the validation of an outcome measure is an ongoing process.

In our first study we evaluated construct validity, internal consistency, factor structure, longitudinal validity and sensitivity to change of the PREE and the pASES-e. We found that the PREE and the pASES-e exhibited acceptable levels of construct validity with the DASH, SF-36 and between themselves; they exhibited high internal consistency with narrow confidence intervals and acceptable levels of sensitivity to change as indicated by large responsiveness indices. The findings of this study corroborated previous studies[3-5] and added additional information about the performance characteristics of these two scales. This study added new information on the internal consistency, factor structure and responsiveness of the PREE and pASES-e.

In our next study (Chapter 3- Objective 2) we synthesized the existing knowledge on the clinical measurement properties of the PREE and the pASES-e by conducting a systematic review. We based our synthesis on nine high-quality studies identified in the peer-reviewed literature. This synthesis indicated that that the PREE and the pASES-e have been shown to demonstrate acceptable levels of psychometric properties in currently published studies. On the other hand, this synthesis also indicated gaps in the literature and included recommendations for future research that were needed to provide more comprehensive evidence about the performance of these clinical measurement tools. The remainder of the thesis was able to address some of these gaps.

Two important and complementary recommendations from the systematic review were the need for more formal content validation of the two PROMs and the need for evaluation with modern psychometric methods like Rasch analysis. Formally analyzing content using an internationally accepted framework like the ICF would help in widespread use of these PROMs across borders, languages and cultures.[6] Since formal content evaluation had not been previously published on either of these measures, this was one important aspect in understanding the conceptual domains addressed by these tools. Conversely, the metrics by which these conceptual domains are measured is equally important. Modern psychometric methods like the Rasch provide analytical techniques to evaluate item performance in terms of targeting, bias and ability to provide linear unbiased interval level scores. [7]

Content analysis is the initial step in validating any patient reported outcome measure. In Chap 4 (objective 3) we used the ICF to analyze the content of the PREE and the pASES-e. We found that both these measures align to the framework of the ICF focusing on the activities and participation section of the ICF indicating that the content of items was well aligned with the proposed domains of pain and upper extremity functional disability. By linking both questionnaires we were able to provide new information about the concepts covered by these two measures which can be useful in comparing their content and future cross-cultural translations. Further we applied recently developed linkage indicators which allowed us to summarize the depth and breadth of the linking between the two measures and the ICF.

Patient centered care and patient oriented research are key themes in current healthcare delivery and research.[9,10] In Chap 5 (Objective 4) we compared the self-endorsed concerns of patients regarding limitations in daily activities to the content covered by the PREE and the pASES-e using ICF as a common ground for comparison. We found both the PREE and the pASES-e covered acceptable ground in terms of patient concerns. When both the PROMs were compared, the PREE covered a larger percentage of concerns then the pASES-e.

Rasch analysis is the new addition to the myriad of psychometric techniques that are available to test the psychometric properties of PROMs.[7] Chapter 6 (Objective 5) of the thesis presented a Rasch analysis of the PREE. The PREE with minimal statistical adjustments and data handling satisfied the Rasch model requirements. This added a substantial body of support for

the rigorous psychometric properties of the PREE, since Rasch analysis had not been employed during development or published in subsequent validation studies.

Integrated Knowledge Translation and clinical implication

This thesis engaged researchers and knowledge users (hand surgeons, hand therapists and physiotherapists) within each phase of the research study. They were central in guiding the development and execution of the research process and thus will engage in sharing information about implementation of these measures in the evaluation of elbow conditions in clinical practice and research. This may facilitate successful knowledge translation of the results.

Clinical implications include the routine use of PREE and the pASES-e in clinics with elbow patients as these two PROMs have better clinical utility with no cost, taking less time (3 minutes) and can be administered and interpreted with minimal training. Also both measures exhibited acceptable levels of psychometric properties which would foster clinician confidence upon the results and interpretations obtained from these two PROMs. Content analysis suggested that the PREE may provide a broader scope of items; although there was insufficient research done within this thesis to suggest that one tool was superior. In fact, given high correlations between the measures one might expect that come to similar conclusions. On the whole both the PREE and the pASES-e can aid clinicians in evaluating, discriminating and predicting pain and disability in patients with elbow disorders.

Contributions of the thesis to measurement theory and practice

Clinical measurement was the backbone of this work and this thesis may contribute substantially to the advancement of the theory and practice of not just measurement of elbow pain and disability but also to the field of clinical measurement in general. In our work we found that the modern psychometric methods provided a different measurement perspective that agreed with the results of classical test methods. For e.g. in the second chapter we found that with the exploratory factor analysis the items of the PREE function section loaded on to multiple factors indicating multidimensionality. During the Rasch analysis the function section was clearly multidimensional and we were able to identify that, the source of multidimensionality was local dependency. Thus the modern methods confirm the findings of the traditional methods and also provide a means of identifying the source of the problem and correcting it. Another new

innovation was the use of linkage indicators to quantify the depth and breadth of linkage between the PROM and the ICF. This is a simple, straightforward and novel method of understanding the linkage and the unique characteristics of the relationship between the items of the PROMs and the ICF. These linkage indicators have not been tested previously. Hence, this needs to be further analyzed in different populations and conditions to understand the full extent of their usefulness and to recommend them for routine use.

Strengths and the limitations

The strengths of this thesis are its wide variety of psychometric approaches that were used to evaluate the PREE and the pASES-e providing evidence from different perspectives and methodologies. We used classical psychometric methods like construct validation, factor analysis to modern psychometric methods like Rasch analysis. Positive findings using multiple statistical methods, that have different assumptions and analytical techniques provides stronger validation and more in-depth psychometric analysis of these two PROMs. This strengthens the clinical applicability by increasing the confidence of the clinician in selecting a PROM. The combination of classical and modern psychometric methods is a particular strength to this thesis.

Beyond understanding elbow disability measures, this thesis also allowed us to expand our findings to the field of clinical measurement. It was interesting to find how the findings from the modern psychometric methods complemented those of classical test methods. Another strength of this thesis is that the individual studies were fully powered. All the three studies that involved human subjects had a sample size of at least 100 subjects. This increases our confidence that the findings of the studies are valid, and allowed for relatively precise estimation of the measurement properties. Another strength of the thesis was that a systematic review was used to identify gaps in the literature that were subsequently addressed in primary research studies.

This thesis also had limitations. Within the context of a thesis, it is not possible to address all of the gaps identified in the literature. Understanding of the measurement properties of tools is an ongoing process that requires tools be evaluated for multiple properties, from multiple perspectives, using qualitative and quantitative methodologies, testing across different patient populations and across different contexts and measurement purposes. One limitation in our work is that the majority of participants came from a tertiary care surgical center. Thus, the

majority of the subjects involved in the studies were patients who had presented for elbow conditions that required surgical management. This would reduce the generalizability of the findings of this thesis to surgical elbow conditions. Future studies should include patients subjected to non surgical treatment as well.

We did address all relevant clinical measurement properties. However we did miss some indices. For example, we did not report the standard error of measurement or minimal clinical important difference for the PREE or pASES-e. Calculation of measurement error and clinically meaningful change would have made these two measures powerful tools in measuring pain and disability arising out of elbow disorders.

Another limitation was that the Rasch analysis of the pASES-e was not performed. Therefore, we cannot inform users whether or not this scale provides interval level scaling or has concerns with respect to targeting or biases tested for the PREE. Given that the content analysis showed that the PREE had a broader coverage of ICF concepts, it would have been useful to determine the extent to which each of these measures demonstrated quantitative psychometric properties assessed by the Rasch model.

Future directions

Future studies should include subjects with elbow disorders that are managed conservatively. This can greatly increase the generalizability of the results obtained. Based on our observation from the systematic review we recommend that studies should explicitly state their hypothesis when conducting psychometric studies; clear, specific and succinct conclusions are also needed. We also recommend the future studies to include anchor based methods to calculate minimal clinically important difference (MCID) and standard error of measurement (SEM). This would greatly help in clinical discrimination. We also recommend the Rasch analysis of PREE at different follow-up time point and population to check the stability of the findings of the current Rasch analysis. Rasch analysis of the pASES-e is highly recommended. Further head to head comparisons of the responsiveness of the PREE and ASES-e and Rasch analyses are needed before definitive conclusions about preference or interchangeability of these measures can be defined with confidence.

REFERENCES

1. Liem IS, Kolling C, Marks M, Nelissen RG, Goldhahn J. Development of a score set to measure function and quality of life in patients suffering from elbow pathology. *Arch Orthop Trauma Surg*. 2012;132(6):831-837.

2. Simmen BR, Angst F, Schwyzer HK, et al. A concept for comprehensively measuring health, function and quality of life following orthopaedic interventions of the upper extremity. *Arch Orthop Trauma Surg*. 2009;129(1):113-118.

3. MacDermid JC. Outcome evaluation in patients with elbow pathology: Issues in instrument development and evaluation. *Journal of Hand Therapy*. 2001;14(2):105-114.

4. Angst F, Goldhahn J, Drerup S, et al. Responsiveness of five outcome measurement instruments in total elbow arthroplasty. *Arthritis Care and Research*. 2012;64(11):1749-1755.

5. Angst F, John M, Pap G, et al. Comprehensive assessment of clinical outcome and quality of life after total elbow arthroplasty. *Arthritis Care and Research*. 2005;53(1):73-82.

6. Cieza A, Stucki G. The international classification of functioning disability and health: Its development process and content validity. *Eur J Phys Rehabil Med*. 2008;44(3):303-313.

7. Andrich D. *Rasch models for measurement.* Vol 07-068. Newbury Park: Sage Publications; 1988:95.

8. The B, Reininga IH, El Moumni M, Eygendaal D. Elbow-specific clinical rating systems: Extent of established validity, reliability, and responsiveness. *J Shoulder Elbow Surg*. 2013;22(10):1380-1394.

9. Haywood K, Marshall S, Fitzpatrick R. Patient participation in the consultation process: A structured review of intervention strategies. *Patient Educ Couns*. 2006;63(1-2):12-23.

10. Small N, Bower P, Chew-Graham CA, Whalley D, Protheroe J. Patient empowerment in long-term conditions: Development and preliminary testing of a new measure. *BMC Health Serv Res*. 2013;13:263-6963-13-263.

APPENDIX 1 ETHICS APPROVAL FORMS

Research Western

Use of Human Participants - Ethics Approval Notice

Principal Investigator: Dr. Joy MacDermid
Review Number: 18021E
Review Level: Delegated
Approved Local Adult Participants: 175
Approved Local Minor Participants: 0
Protocol Title: Psychometric Evaluation of Self-Reported Pain and Disability Measures for Elbow Pathologies
Department & Institution: Surgery, University of Western Ontario
Sponsor:
Ethics Approval Date: May 09, 2011 **Expiry Date:** June 30, 2012
Documents Reviewed & Approved & Documents Received for Information:

Document Name	Comments	Version Date
UWO Protocol		

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the UWO Updated Approval Request Form.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

The Chair of the HSREB is Dr. Joseph Gilbert. The UWO HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

[redacted]

Ethics Officer to Contact for Further Information

Janice Sutherland [redacted]	Grace Kelly [redacted]

This is an official document. Please retain the original in your files.

The University *of* Western Ontario
Office of Research Ethics

[redacted]

LAWSON HEALTH RESEARCH INSTITUTE

FINAL APPROVAL NOTICE

RESEARCH OFFICE REVIEW NO.: R-11-209

PROJECT TITLE: Psychometric Evaluation of Self-Reported Pain and Disability Measures for Elbow Pathologies

PRINCIPAL INVESTIGATOR: Dr. Joy Christine MacDermid

DATE OF REVIEW BY CRIC: May 15, 2011

Health Sciences REB#: 18021E

Please be advised that the above project was reviewed by the Clinical Research Impact Committee and the project:

Was Approved

PLEASE INFORM THE APPROPRIATE NURSING UNITS, LABORATORIES, ETC. BEFORE STARTING THIS PROTOCOL. THE RESEARCH OFFICE NUMBER MUST BE USED WHEN COMMUNICATING WITH THESE AREAS.

Dr. David Hill
V.P. Research
Lawson Health Research Institute

[redacted]

cc: Administration

Western Research

Research Ethics

Use of Human Participants - Initial Ethics Approval Notice

Principal Investigator: Dr. Joy MacDermid
File Number:104647
Review Level:Delegated
Protocol Title:Linking of two elbow Self-report pain and disability measures to the ICF
Department & Institution:Schulich School of Medicine and Dentistry\Surgery,Western University
Sponsor:
Ethics Approval Date:December 20, 2013 Expiry Date:May 25, 2014
Documents Reviewed & Approved & Documents Received for Information:

Document Name	Comments	Version Date
Instruments	Patient Specific Functional Scale	
Western University Protocol		
Letter of Information & Consent	Modified letter of information	2013/12/04

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the University of Western Ontario Updated Approval Request Form.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

The Chair of the HSREB is Dr. Joseph Gilbert. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

[redacted]

for Further Information

Erika Basile	Grace Kelly	Mina Mekhail	Vicki Tran
[redacted]	[redacted]	[redacted]	[redacted]

This is an official document. Please retain the original in your files.

[redacted]

LAWSON
HEALTH RESEARCH INSTITUTE

LAWSON FINAL APPROVAL NOTICE

LAWSON APPROVAL NUMBER: R-13-521

PROJECT TITLE: Linking of two elbow Self-report pain and disability measures to the ICF

PRINCIPAL INVESTIGATOR: Dr. Joy MacDermid

LAWSON APPROVAL DATE: January 15, 2013

Health Sciences REB#: 104647

Please be advised that the above project was reviewed by the Clinical Research Impact Committee and Lawson Administration and the project:

Was Approved

Please inform the appropriate nursing units, laboratories, etc. before starting this protocol. The Lawson Approval Number must be used when communicating with these areas.

Dr. David Hill
V.P. Research
Lawson Health Research Institute

[redacted]

cc: Administration

Western

Research Ethics

Use of Human Participants - Ethics Approval Notice

Principal Investigator: Dr. Joy MacDermid
File Number:5697
Review Level: Delegated
Approved Local Adult Participants:2400
Approved Local Minor Participants:0
Protocol Title: Wrist and Elbow Outcomes Measures Database -15602E
Department & Institution: Schulich School of Medicine and Dentistry\Surgery,Western University
Sponsor:
Ethics Approval Date: October 29, 2012 **Expiry Date:**August 31, 2014
Documents Reviewed & Approved & Documents Received for Information:

Document Name	Comments	Version Date
Revised Study End Date	The study end date has been extended to August 31, 2014.	
Increase in number of local Participants	The number of study participants has been increased to 2400 (120 per year times 20 years).	

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the University of Western Ontario Updated Approval Request Form.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

The Chair of the HSREB is Dr. Joseph Gilbert. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

[redacted]

Ethics Officer to Contact for Further Information

Janice Sutherland	Grace Kelly	Shantel Walcott
[redacted]	[redacted]	[redacted]

This is an official document. Please retain the original in your files.

[redacted]

APPENDIX 2: LETTER OF INFORMATION

Dr. Joy C. MacDermid,
Principal Investigator
Joshua Israel Vincent, PhD Student
Health and Rehabilitation Sciences,
Faculty of Health Sciences,
University of Western Ontario, London,
Canada

Address:
[redacted]

PARTICIPANT INFORMATION SHEET

Title of Project: Linking of two elbow Self-report pain and disability measures to the ICF

Principal Investigator: Dr. Joy MacDermid, PhD, MSc, BScPT, BSc, Professor, Health and Rehabilitation Sciences, Faculty of Health Sciences, University of Western Ontario, London, Ontario, Canada

Co-Investigator: Joshua Israel Vincent, MPT, BPT, PhD Student, Health and Rehabilitation Sciences, Faculty of Health Sciences, University of Western Ontario, London, Ontario, Canada

Co-Investigator: Dr.Ruby Grewal, MD, MSc, FRCSC, Asst. Professor, Department of surgery, University of Western Ontario, London, Ontario, Canada and Hand Surgeon, Hand and Upper Limb Center, St. Joseph's Hospital, London, Ontario, Canada

Co-Investigator: Dr. Graham King, MD, MSc, FRCSC, Professor, Department of Surgery, University of Western Ontario, London, Ontario, Canada and Director, Hand and Upper Limb Center, St. Joseph's Hospital, London, Ontario, Canada

Study ID #

You are invited to be in a research study conducted by Joshua Israel Vincent involving collection of information on functional activities that are important to you through a questionnaire. You are invited to participate in this study because you have problems with your elbow joint and you match the inclusion criteria of this study.

The purpose of this letter of information is to provide you with detailed information about the research study, and for you to understand what is involved and the potential risks and benefits in order to decide whether or not you want to be a part of this research study. Once you understand the study, you will be asked to sign this form if you wish to participate. Please take your time to make your decision. Feel free to discuss it with your friends and family, or your family physician.

What is the purpose of this study? The purpose of the study is to see if the content of the commonly used elbow pain and disability questionnaires can be linked to internationally recognized World health organization's (WHO) classification system called the International Classification of Functioning Disability and health (ICF). The secondary purpose is to compare and see if these questionnaires cover things that are important to patients.

Eligibility criteria: Individuals aged between 20 and 80 years with an elbow problem, that can read and write English can participate in this study. Individuals with communication problems and difficulties in comprehension or with a history of cancer will not be able to participate in this study.

What am I expected to do? If you agree to participate, you will be asked to fill in a form called as 'Patient Specific Functional Scale'. You will be asked to identify up to five important activities that you are unable to do or are having difficulty with as a result of your elbow problem. Then you will be asked to rate the level of difficulty for each of the activity on a scale 0-10 where 0 being unable to do the activity and 10 being able to perform at the same level as before injury. It is anticipated that the entire task will take approximately 10 minutes. You will be asked to fill this form here at the hand and Upper limb clinic. There are no follow-ups for this study. No extra visits are required. There will be a total of 100 participants. When you come in for the treatment for your elbow problem the standard care you will be receiving includes routine assessment of your elbow disorder and based on that assessment, you would receive appropriate medical or surgical care. This study in no way would alter the standard care that you would receive

Is there any possible risks/harm? There are no known or anticipated risks or discomforts associated with participating in this study as it involves only filling in a questionnaire.

What are the possible benefits? You may not directly benefit from participating in this study but information gathered may provide benefits to society as a whole.

Will I be paid to participate in this study? No. However, you will be provided parking passes to cover the incidental costs of parking to participate in this study. There is no costs that are involved on your part.

Voluntary Participation: Participation in this study is voluntary. You may refuse to participate, refuse to answer any questions or withdraw from the study at any time with no effect on your future care at St. Joseph's Healthcare.

Confidentiality: All data collected will remain confidential and accessible only by the investigators of this study. All data collected will remain anonymous. If the results are published, your name will

not be used. If you choose to withdraw from this study, your data will be removed and destroyed from our database. Our research records will be stored in the following manner: locked in a cabinet in a secure office; audio and video recordings will be reviewed only by members of the research team and they will be destroyed after 10 years. If we find information we are required by law to disclose, we cannot guarantee confidentiality. Representatives of The University of Western Ontario Health Sciences Research Ethics Board may contact you or require access to your study-related records to monitor the conduct of the research.

Will I be provided with the results of the study? If the results of the study are published, your name will not be used. If you would like to receive a copy of any potential study results, please provide your name and contact number on a piece of paper separate from the Consent Form

Questions about the Research? If you have any questions about your rights as a research participant or the conduct of this study, you may contact The Office of Research Ethics [redacted] email: [redacted] Dr. David Hill, Scientific Director, Lawson Health Research Institute at [redacted] Dr. Joy MacDermid the principal investigator of this study at [redacted] Email: [redacted]

APPENDIX 3 CONSENT STATEMENT

TITLE: Linking of two elbow Self-report pain and disability measures to the ICF

SIGNATURE OF RESEARCH PARTICIPANT:

I have read the Letter of Information, (have had the nature of the study explained to me) and I agree to participate. I have had the opportunity to ask questions, all questions have been answered to my satisfaction.

_____ _____

Participant's Signature Date

Consent form administered and explained in person by: _____

_____ _____

Signature Date

SIGNATURE OF INVESTIGATOR:

In my judgment, the participant is voluntarily and knowingly giving informed consent and possesses the legal capacity to give informed consent to participate in this research study.

_____ _____

Signature of Investigator Date

APPENDIX 4 COPYRIGHT PERMISSION

Gmail by Google

[redacted]

RE: permission for doctoral thesis

[redacted] Mon, Aug 11, 2014 at 4:35 PM

Hi Joshua,

Thank you for contacting us about this matter. As an author on this paper, you are free to use your own work in your thesis. When using/citing material from this article, please use the following format:

Reproduced with permission from Vincent JI, MacDermid JC, King GJ, Grewal R. Validity and sensitivity to change of patient-reported pain and disability measures for elbow pathologies. *J Orthop Sports Phys Ther.* 2013;43:263-274. http://dx.doi.org/10.2519/jospt.2013.4029. Copyright ©*Journal of Orthopaedic & Sports Physical Therapy*®

However, please note that the Appendix in that article is copyrighted to Dr MacDermid, so you need to obtain additional permission from her to use that portion of the article.

Please let me know if you have further questions.

Kind regards,

Corey

Corey Parker

Copy Editor

Journal of Orthopaedic & Sports Physical Therapy

[redacted]

[redacted]
Sent: Saturday, August 09, 2014 2:11 PM
To: Anthony Willard
Subject: Re: JOSPT Author Proofs

Hi Anthony,

I am a University of Western Ontario graduate student completing my Doctoral thesis entitled "Psychometric evaluation of two pain and disability self-report measures for elbow disorders".

I would like permission to allow inclusion of the following material in my thesis**: Validity and Sensitivity to Change of Patient-Reported Pain and Disability Measures for Elbow Pathologies** published in the Journal of Orthopaedic & Sports Physical Therapy. Volume 43, number 4, April 2013

My thesis will be available in full-text on the internet for reference, study and / or copy. Except in situations where a thesis is under embargo or restriction, the electronic version will be accessible through the Western Libraries web pages, the Library's web catalogue, and also through web search engines. I will also be granting Library and Archives Canada and ProQuest/UMI a non-exclusive license to reproduce, loan, distribute, or sell single copies of my thesis by any means and in any form or format. These rights will in no way restrict republication of the material in any other form by you or by others authorized by you.

The material will be attributed through a citation.

Please confirm in writing or by email that these arrangements meet with your approval.

Sincerely

Joshua Vincent

On Tue, Apr 9, 2013 at 2:18 PM, [redacted]

Dear Mr. Vincent,

Enclosed with this email is a final PDF of your Research Report, *Validity and Sensitivity to Change of Patient-Reported Pain and Disability Measures for Elbow Pathologies* published in the April 2013 issue of the *JOSPT*. This file is an electronic reprint for authors to archive and distribute as they please. We ask in return that they include the following statement with distributed copies: "Reprinted with permission of the *Journal of Orthopaedic and Sports Physical Therapy*, and the Sports Physical Therapy Section and the Orthopaedic Section of the American Physical Therapy Association."

Please remember that each author is entitled to a complimentary paper copy of the April 2013 issue as well. If you wish to receive your copy, please email me at [redacted]

Thank you for your contribution to the *JOSPT*.

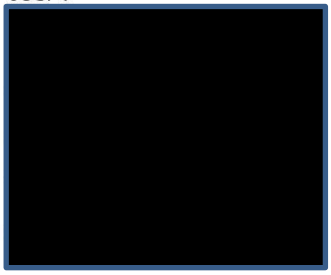
Best regards,

Anthony

Anthony Willard

Manuscript & Administrative Coordinator

JOSPT



CURRICULUM VITAE

EDUCATION:

- ***Bachelor of Physiotherapy (2007)*** *Tamil Nadu Dr. MGR Medical University, Chennai, India*
- ***Master of Physiotherapy (2009)*** *Tamil Nadu Dr. MGR Medical University, Chennai, India*
- ***Ph.D. (Expected Sept 2014)*** *University of Western Ontario, London, ON*

PROFESSIONAL EMPLOYMENT

- ***Designation:*** *Research Assistant* ***(Sept 2012- till date)***
  Employer: *Roth-MacFarlane Hand and upper limb, St. Joseph's healthcare London, London, ON.*

- ***Designation:*** *Teaching Assistant* ***(Jan 2012-Apr 2012; Jan 2011-Apr 2011; Jan 2013-Apr 2013)***
  Employer: *School of physiotherapy, University of Western Ontario, London, ON*

- ***Designation:*** *Research Assistant (****Sept 2011- Dec 2011)***
  Employer: *International Center for Health Innovation, Richard Ivey School Of Business, University of Western Ontario, London, ON.*

- ***Designation:*** *Tutor in Physical therapy* ***(Jan 2010- July 2010)***
  Employer: *Pondicherry Institute of Medical Sciences, Pondicherry, India*

- ***Designation:*** *Consultant Physiotherapist (****Nov 2007- Jan 2010)***
  Employer: *Grace Physiotherapy Center, Chennai, India*

AWARDS

- **"CIHR - Institute of Community Support (ICS) travel Award"** (2013) Awarded by the Institute of Musculoskeletal Health and Arthritis (CIHR). National competition. Amount $ 1000

- **Ontario Graduate Scholarship (OGS)** (2013-2014) Awarded by the Ministry of Training, College and Universities' (MTCU). Provincial competition. Amount $ 15,000

- **"CIHR - Institute of Community Support (ICS) travel Award"** (2012) Awarded by the Institute of Gender and Health (CIHR). National competition. Amount $ 2433

- **"CIHR - Institute of Community Support (ICS) travel Award"** (2011) Awarded by the Institute of Musculoskeletal Health and Arthritis (CIHR). National competition. Amount $ 1000

- **"Best Scientific Paper award" (Nominated)** at the American Society of Hand Therapists 37th Annual Meeting, Nashville, TN: Psychometric Evaluation of Self-reported Pain and Disability Measures for Elbow Pathologies (2011)
- **"Best outgoing student of the class of 2007 award"** (for highest academic achievement in the graduating class-Master of Physiotherapy) awarded by Vel's College of Physiotherapy, Chennai, India.(2009)
- **"Best outgoing student of the class of 2002 award"** (for highest academic achievement in the graduating class-Bachelor of Physiotherapy) awarded by Vel's College of Physiotherapy, Chennai, India.(2007)
- **"Best Scientific Paper award"** at Entrevue Physios-07: Muscle Energy Techniques (METS). Pondicherry, India (2007)
- **"Best Scientific Paper award"** at Physiofest 07: The healing energy. Chennai, India.(2007)
- **"Best Scientific Paper award"** at Physiofest 06: Disentangling the knots. Chennai, India.(2006)

PEER REVIEWED PUBLICATIONS (Total 9)

- ***Vincent, JI,*** *MacDermid*, JC. (2014). The Patient-Rated Tennis Elbow Evaluation (PRTEE). Journal of physiotherapy, (Accepted In-Press).
- ***Vincent, JI,*** MacDermid JC, Michlovitz SL, Rafuse R, Wells-Rowswell C, Wong O, Bisbee L. (2014) The Push off test: Development of simple reliable test of Upper Extremity Weight-Bearing Capability. Journal of Hand Therapy. (In-press)
- ***Vincent, JI,*** MacDermid, JC, Grewal, R., Balachandran, D., Sekar, V.P. (2014) Translation of Oswestry Disability Index into Tamil with Cross Cultural Adaptation and Evaluation of Validity, Reliability and Responsiveness. Open Orthopedics 8:11-19
- ***Vincent, JI,*** *MacDermid* JC, King GJW, Grewal R (2013) Response: Letter to the Editor-in-Chief "Critical Assessment of Patient-Reported Outcome Measures" The Journal of orthopaedic and sports physical therapy, 2013;43(7):513-514.
- ***Vincent, JI,*** Vandervoort, A. A., & MacDermid, JC. (2013). A Literature Synthesis Indicates Very Low Quality, but Consistent Evidence of Improvements in Function after Surgical Interventions for Primary Osteoarthritis of the Elbow. Arthritis, 2013.
- ***Vincent, JI.*** MacDermid, JC, King, G. J., Grewal, R. (2013). Validity and sensitivity to change of patient-reported pain and disability measures for elbow pathologies. The Journal of orthopaedic and sports physical therapy, 43(4), 263-274.

- ***Vincent, JI,*** MacDermid, JC (2012). The Patient-Rated Elbow Evaluation (PREE). Journal of physiotherapy, 58(4), 274.

- MacDermid, JC, **Vincent, JI**, Kieffer, L., Kieffer, A., Demaiter, J., & MacIntosh, S. (2012). A Survey of Practice Patterns for Rehabilitation Post Elbow Fracture. The open orthopaedics journal 6: 429 - 439.

- MacDermid, JC, **Vincent, JI,** Gan, BS, & Grewal, R. (2012). A blinded placebo-controlled randomized trial on the use of Astaxanthin as an adjunct to splinting in the treatment of carpal tunnel syndrome. Hand, 7(1), 1-9.

Submitted (Total 4)

- **Vincent, JI,** MacDermid, JC, King GJW, Grewal, R. Linking of the Patient Rated Elbow Evaluation (PREE) and the American Shoulder and Elbow Surgeons – Elbow questionnaire (pASES-e) to the International Classification of Functioning Disability and Health (ICF) and Hand Core Sets. Journal of Hand Therapy (2014)

- **Vincent, JI**, MacDermid, JC, King GJW, Grewal, R. Measurement of elbow Pain and Function: a systematic review of the psychometric properties of two elbow self-report measures. Arthritis Care & Research. (2014)

- MacDermid, JC, Arumugam, A., **Vincent, JI**, Payne, KL, SO, AK. Reliability of Three Landmarking Methods for Dual Inclinometry Measurements of Lumbar Flexion and Extension. Spine

- MacDermid, JC, Arumugam, A., **Vincent, JI,** Carol, KL. The Reliability and validity of the Computerised Double Inclinometer in Measuring Lumbar Mobility. Open orthopedics

Invited presentations (Total 2)

- ***Vincent JI,**** Outcome measures used in the management of elbow instability. 9th International Federation of Societies of Hand Therapy triennial congress, March 4th -8th, 2013 in New Delhi, India. (International)

- ***Vincent JI,**** Rasch analysis of the Patient Rated Elbow Evaluation (PREE) Health and Rehabilitation Sciences common seminar series, University of Western Ontario, London ON (Institutional)

Selected published peer-reviewed abstracts from conference proceedings (Total 9)

- **Vincent JI*,** MacDermid JC., King GJ., Grewal R., Baseline predictors of self-reported pain and disability after 2 years in patients who underwent Biceps tendon repair. ASHT 39th Annual meeting. Journal of Hand Therapy 2014; 27(3):e7 – e8.

- **Vincent JI***, MacDermid JC, Grewal R., Balachandran D., Sekar V.P. Translation of Oswestry disability index into Tamil with cross cultural adaptation and evaluation of reliability and validity. 20th annual conference of the international society for quality of life research. *Quality of Life Research*. 2013; 22(1):60.

- **Vincent JI*,** MacDermid JC, Grewal R. The evaluation of the patient rated elbow evaluation using Rasch analysis. 19th annual conference of the international society for quality of life research. *Quality of Life Research*. 2012; 21 (Suppl):39.

- **Vincent JI,** MacDermid JC, Michlovitz SL*. Psychometric testing of a clinical test to assess upper extremity weight-bearing capability: The push-off test. ASHT 38th annual meeting. *Journal of Hand Therapy*. 2012; 25(4):e4-e4.

- **Vincent JI***, MacDermid JC, King G, Grewal R. Psychometric evaluation of self-reported pain and disability measures for elbow pathologies. ASHT 37th annual meeting. *Journal of Hand Therapy*. 2011; 24(4):381-382.

- **Vincent JI*,** MacDermid, JC, King, GJW, Grewal, R. Psychometric evaluation of self-reported pain and disability measures in elbow pathologies. 18th annual conference of the international society for quality of life research. *Quality of Life Research*. 2012;20(Suppl 1):35-36

Selected Conference Presentations (* Presenting author) (Total 13)

- **Vincent JI,*** MacDermid JC., King GJW, Grewal R., (2013)Baseline predictors of self-reported pain and disability after 2 years in patients who underwent Biceps tendon repair. ASHT 39th Annual meeting Oct 2013, Chicago, IL USA (International)

- **Vincent JI,*** MacDermid JC, Grewal R, Balachandran D., Sekar VP., (2013) Translation of Oswestry disability index into Tamil with cross cultural adaptation and evaluation of reliability and validity. International Society of Quality of Life 20th Annual conference, Oct 2013, Miami, FL USA (International)

- **Vincent JI,*** MacDermid, JC, King, G.J., Grewal, R. (2013) Linking of Two Elbow Pain and Disability Self-Report Measures to the ICF. 9th International Federation of Societies for Hand Therapy (IFSHT) Triennial Congress. Mar 2013, New Delhi, India (International)

- **Vincent JI,*** MacDermid JC, Grewal R. The evaluation of the patient rated elbow evaluation using Rasch analysis. International society of Quality of Life, 18th Annual Conference, Budapest, Oct 2012 (International)

- **Vincent JI,** MacDermid JC, Michlovitz SL*. Psychometric testing of a clinical test to assess upper extremity weight-bearing capability: The push-off test. The American Society of Hand Therapists 38th Annual Meeting, San Diego. Oct 2012 (International)

- MacDermid JC, **Vincent, JI,** * Kieffer L, Kieffer A, Demaiter J, MacIntosh S. (2012) A survey of practice patterns for rehabilitation post elbow fracture. 2012 Combined Meeting of the Canadian Society of Hand Therapists and the Canadian Society for Surgery of the Hand, Toronto, ON. May 2012. (National)

- **Vincent, JI,*** MacDermid, JC, King, G.J., Grewal, R. (2011) Evaluation of Validity, responsiveness and factor structure of Pain and disability measures for elbow disorders. International society of Quality of Life, 17th Annual Conference, Denver, CO. Oct, 2011(International)

- **Vincent, JI,*** MacDermid, JC, King, G.J., Grewal, R. (2011) Psychometric Evaluation of Self-reported Pain and Disability Measures for Elbow Pathologies. The American Society of Hand Therapists 37th Annual Meeting, Nashville, TN. Sept, 2011 (International) (Nominated for the best scientific paper award)

- **Vincent, JI,*** MacDermid, JC, King, G.J., Grewal, R. (2011) Psychometric Evaluation of Self-reported Pain and Disability Measures for Elbow Pathologies. 2012 Combined Meeting of the Canadian Society of Hand Therapists and the Canadian Society for Surgery of the Hand, Vancouver, B.C. April, 2011. (National)

- **Vincent, JI,*** (2007) The healing energy. "ENTREVUE PHYSIOS-07" conducted by College of Physiotherapy, Mother Theresa Institute of Health Sciences, Pondicherry, India. Nov, 2007. (National) (Best Scientific paper award)

- **Vincent, JI,*** (2007) The Muscle Energy Techniques. Vel's College of Physiotherapy, Chennai, India. Dec, 2007. (Institutional) (Best Scientific paper award)

- **Vincent, JI,*** (2006) Disentangling the Knots. "ENTREVUE PHYSIOS-06" conducted by College of Physiotherapy, Mother Theresa Institute of Health Sciences, Pondicherry, India. Oct, 2006. (National)

- **Vincent, JI,*** (2006) The Myofascial Trigger Point Release. Vel's College of Physiotherapy, Chennai, India. Dec, 2006. (Institutional) (Best Scientific paper award)

PROFESSIONAL AND RESEARCH ACTIVITIES:

- ***Reviewer,*** *Arthritis , Journal of Hand Therapy,*
- ***Field Mentor (2013-14)*** *Physiotherapy Field at The HRS dept, University of Western Ontario*
- ***Member,*** *Tennis elbow Clinical Practice Guidelines (CPG) development group*
- ***Member,*** *Psychometric special interest group*
- ***Member,*** *Cross-cultural adaptation special interest group*
- ***Member,*** *Response-shift special interest group*
- ***Member,*** *International Society for Quality of Life Research (ISOQOL)*
- ***Member,*** *American Society of Hand Therapists (ASHT)*
- ***Member,*** *International Federation for the Societies of Hand Therapists (IFSHT)*
- ***Member,*** *World Confederation Of Physical Therapy (WCPT)*
- ***Member,*** *The Indian Association of Physiotherapists (IAP)*
- ***Certified First Aider (CPR-AED)-*** *Canadian Red Cross*
- ***Trained Teaching Assistant-*** *Teaching support center of The University of Western Ontario*
- ***Certified Bone Safety Evaluator -*** *Certified by Ion med*

SUPERVISORY EXPERIENCE

- Boake, B.; Childs, T; Soules, T.; (2014) Rasch analysis of the SPADI. (Co-supervisor in Master's in Physiotherapy research project)
- Mariam (Sept 2013 – Jun 2014) Distal Radius Fractures and the Potential Implications. (High school co-op student)
- Suski, K. (Sept 2012 – Jan 2013) Distal Radius Fractures and the Potential Implications. (High school co-op student)

- Zeb, R. (Oct 2011 – Jun 2012) Analysis using Dartfish. (High school co-op student)

VOLUNTEERING ACTIVITIES

- *Volunteering at the Physiotherapy department of St. Joseph's hospital, London, ON*
- *Volunteering with "Helping Hands ", a not-for-profit organisation distributing lunches in downtown London*

References provided upon request.